MOLECULAR ECOLOGY

A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (*Salvelinus namaycush*)

FREDERICK GOETZ,* DANIEL ROSAUER,* SHAWN SITAR,† GILES GOETZ,* CRYSTAL SIMCHICK,* STEVEN ROBERTS,‡ RONALD JOHNSON,§ CHERYL MURPHY,¶ CHARLES R. BRONTE** and SIMON MACKENZIE††

*Great Lakes WATER Institute, University of Wisconsin-Milwaukee, 600 E. Greenfield Ave., Milwaukee, WI 53204, USA, †Marquette Fisheries Research Station, Michigan Department of Natural Resources, 484 Cherry Creek Rd., Marquette, MI 49855, USA, ‡School of Aquatic and Fishery Sciences, University of Washington-Seattle, 1122 NE Boat Street, Seattle, WA 98105, USA, §NOAA Fisheries Service, Northwest Fisheries Science Center, 2725 Montlake Blvd. East, Seattle, WA 98112-2097, USA, ¶Department of Fisheries and Wildlife, Lyman Briggs College, Michigan State University, 13 Natural Resources Building, East Lansing, MI 48824-1221, USA, **U.S. Fish and Wildlife Service, Green Bay Fish and Wildlife Conservation Office, 2661 Scott Tower Drive, New Franken, WI 54229, USA, ††Departament de Biología Cellular, Fisiologia i Immunologia, Unitat de Fisiologia Animal, Facultat de Biociencies, Universitat Autonoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

Abstract

In Lake Superior there are three principal forms of lake trout (*Salvelinus namaycush*): lean, siscowet and humper. Wild lean and siscowet differ in the shape and relative size of the head, size of the fins, location and size of the eyes, caudal peduncle shape and lipid content of the musculature. To investigate the basis for these phenotypic differences, lean and siscowet lake trout, derived from gametes of wild populations in Lake Superior, were reared communally under identical environmental conditions for 2.5 years. Fish were analysed for growth, morphometry and lipid content, and differences in liver transcriptomics were investigated using Roche 454 GS-FLX pyrosequencing. The results demonstrate that key phenotypic differences between wild lean and siscowet lake trout such as condition factor, morphometry and lipid levels, persist in these two forms when reared in the laboratory under identical environmental conditions. This strongly suggests that these differences are genetic and not a result of environmental plasticity. Transcriptomic analysis involving the comparison of hepatic gene frequencies (RNA-seq) and expression (quantitative reverse transcription–polymerase chain reaction (qPCR)) between the two lake trout forms, indicated two primary gene groups that were differentially expressed; those involving lipid synthesis, metabolism and transport (acyl-CoA desaturase, acyl-CoA binding protein, peroxisome proliferator-activated receptor gamma, and apolipoproteins), and those involved with immunity (complement component C3, proteasome, FK506 binding protein 5 and C1q proteins). The results demonstrate that RNA-seq can be used to identify differentially expressed genes; however, some discrepancies between RNA-seq analysis and qPCR indicate that methods for deep sequencing may need to be refined and/or different RNA-seq platforms utilized.

Keywords: Lake Superior, lake trout, lean, phenotypic differentiation, pyrosequencing, siscowet, transcriptomics

Received 30 June 2009; revision received 15 September 2009; accepted 28 September 2009

Correspondence: Frederick Goetz, Fax: +1 414 382 1705;
E-mail: rick@uwm.edu

Introduction

Resource polymorphisms have been recorded in a diverse number of vertebrates from fish to mammals and have been instrumental in addressing a variety of questions in competition, phenotypic plasticity, speciation and niche utilization (see, Skulason & Smith 1995; for review). A number of fish species contain sympatric populations exhibiting distinct ecological, morphological and genetic differences. They are most common in the Salmonidae (trout, salmon, whitefish, char), Gasterosteidae (sticklebacks) and Osmeridae (smelts) but include fish in other families as well (Taylor 1999). These fish have been the focus of many studies, particularly related to speciation (Jonsson & Jonsson 2001; McKinnon & Rundle 2002).

Lakes in the Northern Hemisphere, formed as a result of the last glaciation event, contain a number of fish ecotypes representing several fish taxa. In many of these examples, sympatric populations occupy benthic and limnetic habitats and involve phenotypes specialized for resources and feeding in these habitats (Robinson & Wilson 1994; Skulason & Smith 1995). A large number of these ecotypes fall within the Salmonidae including the chars that exhibit enormous phenotypic diversity, especially in lake trout (*Salvelinus namaycush*) and to an even greater extent in Arctic char (*Salvelinus alpinus*). Historically, at least 15 separate subspecies of Arctic char were recognized (Adams & Maitland 2006). There is a high degree of genetic differentiation as determined by microsatellites among Arctic char populations across many lakes in Europe and among as well as across populations within the same lake or body of water (Wilson *et al.* 2004; Adams *et al.* 2006). This suggests that reproductive isolation of sympatric Arctic char populations is responsible for maintaining different forms.

Similarly, various lake trout forms have been reported in the Laurentian Great Lakes (Agassiz 1850; Brown *et al.* 1981; Goodier 1981) and more recently from large Canadian shield lakes (Blackie *et al.* 2003; Alfonso 2004; Zimmerman *et al.* 2007). Overharvesting and predation by sea lamprey were responsible for the collapse of lake trout stocks across the Great Lakes and the loss of diversity in all the lakes with the exception of Lake Superior (Krueger & Ihssen 1995; Hansen 1999). This remnant diversity is now used or being considered for reintroduction into the other Great Lakes (Bronte *et al.* 2008; Markham *et al.* 2008). In Lake Superior the three principal lake trout forms are the lean, siscowet and humper lake trout (Lawrie & Rahrer 1973; Moore & Bronte 2001). The origins of the three forms are not entirely clear. Based on mitochondrial DNA analysis, it has been suggested that siscowets evolved within the last 14 000 years from mixed origins (Wilson & Mandrak 2004). Microsatellite

analysis of Lake Superior lake trout supports the divergence of morphotypes prior to the partitioning of proglacial lakes into the Great Lakes (Page *et al.* 2004). However, an alternative hypothesis proposes that the evolution of the different forms occurred during or after the formation of proglacial lakes and that the humper diverged from lean lake trout first, followed by the siscowet (Eshenroder 2008). This would make the origins of siscowet postglacial. The shape and relative size of the head, size of the fins, location and size of the eyes and caudal peduncle shape can visually differentiate these forms. Lean, siscowet and humper lake trout have been experimentally delineated on the basis of osteology (Agassiz 1850; Burnham-Curtis & Smith 1994), morphology (Khan & Qadri 1970; Burnham-Curtis 1993; Moore & Bronte 2001) as well as microsatellite analysis (Page *et al.* 2004). Besides the differences in morphology, siscowet and lean lake trout occupy different habitats in Lake Superior. Lean lake trout occupy shallower nearshore waters at depths less than 100 m, whereas siscowet lake trout are most abundant at depths greater than 100 m (Bronte *et al.* 2003), including the deepest parts of Lake Superior (∼400 m; Sitar *et al.* 2008). The diets of these forms also differ as a result of their depth distribution: lean lake trout feed mostly on rainbow smelt (*Osmerus mordax*), cisco (*Coregonus artedii*) and slimy sculpin (*Cottus cognatus*), whereas siscowet lake trout in deepwater, feed on coregonines (*Coregonus* sp.), deepwater sculpin (*Myoxocephalus thompsoni*) and burbot (*Lota lota*) (Conner *et al.* 1993; Harvey *et al.* 2003; Ray *et al.* 2007; Sitar *et al.* 2008). Siscowet also have a significantly higher percentage of skeletal muscle lipid (Eschmeyer & Phillips 1965). Lean lake trout have been the primary focus of restoration efforts, harvest management and research; however, siscowet lake trout currently make-up most of the lake trout biomass in Lake Superior, as they probably did historically (Bronte *et al.* 2003).

Distinct lake trout forms also occur in other large lakes in North America including Great Bear Lake (Blackie *et al.* 2003; Alfonso 2004), Great Slave Lake (Zimmerman *et al.* 2006) and Lake Mistassini (Zimmerman *et al.* 2007). A deepwater form similar to siscowet and a shallower lean-like form have been reported from Great Slave Lake (Zimmerman *et al.* 2006), and a humper form was described from Lake Mistassini (Zimmerman *et al.* 2007). Forms from deep and shallow habitats from these lakes are differentiated by body shape (deep vs. elongate), length of the pectoral fins (long vs. short) and buoyancy (high vs. low) (Zimmerman *et al.* 2006). Differences in these characters are thought to be related to the energetics of swimming and movement in the water column (Zimmerman *et al.* 2006).

Page *et al.* (2004) reported genetic differences between lean and siscowet lake trout, but the selectively neutral

markers used may be unrelated to the actual phenotypic variation observed. Theoretically, phenotypic variation could be a result of genetic regulation and/or environmental plasticity. For example, the different habitats occupied by these lake trout forms vary greatly in pressure, light, temperature and diet and could be responsible for some phenotypic variation. A number of studies have addressed the genetic or environmental basis for phenotypic variation in Arctic char by rearing populations of different morphs under identical environmental conditions in the laboratory (Nordeng 1983; Svedang 1990; Hindar & Jonsson 1993; Skulason *et al.* 1996; Klemetsen *et al.* 2002; Adams & Huntingford 2004). The results have varied, indicating in some cases predominantly genetic control (Skulason *et al.* 1996) to predominantly environmental control (Adams & Huntingford 2004). It has been suggested that one reason for the differing emphasis on genetic vs. environmental control in these studies is that different polymorphisms are at different stages of divergence (Adams & Huntingford 2004). Surprisingly, there have been few attempts to look at the control of phenotypic differences between lean and siscowet lake trout. A comparison of lean and siscowet lake trout raised under similar environmental conditions was conducted at the Marquette State Fish Hatchery (Marquette, MI) but only results on lipid levels were published (Eschmeyer & Phillips 1965). The results indicated that cultured siscowet lake trout retained the same high lipid levels observed in wild siscowet, and that hybrids between siscowet and lean lake trout had intermediate lipid levels, which strongly suggested that lipid content was genetically determined. Another study conducted by the Michigan Department of Natural Resources at the Thompson State Hatchery, cross-bred siscowet, humper and lean lake trout (Stauffer & Peck 1981). The initial results indicated differences in some meristic characters measured, 'but none were definitive' (Stauffer & Peck 1981). However, Burnham-Curtis (1993) reanalysed the Stauffer & Peck (1981) data using principal component analysis and concluded that morphological and meristic differences observed in the progeny had some genetic basis and were not exclusively environmental. However, some caution must be applied because the fish analysed came from crosses of only two males and two females per lake trout form.

While the results of these two studies indicate a genetic basis for some differences observed between wild lean and siscowet lake trout, there has not been a comprehensive study that investigates a wide range of phenotypic differences between these forms within the same individuals reared under tightly controlled environmental conditions. Thus, one objective of the present study was to analyse growth, condition factor, morphometry and lipid content between lean and siscowet lake trout reared under identical conditions. Based on the results of the laboratory rearing studies on Arctic char ecotypes, we hypothesized that some of the phenotypic traits observed in wild lean and siscowet lake trout would be maintained under laboratory rearing conditions, but some would not (i.e. genetic and environmental contributions). Alternatively, the magnitude of the differences in the expression of certain phenotypes between laboratory-reared lake trout forms would be different in comparison to wild lean and siscowet lake trout.

There has been a strong effort recently to utilize global genomic approaches such as DNA microarrays, to determine the basis for phenotypic divergence and parallel evolution in fish ecotypes (Goetz & MacKenzie 2008). Microarrays have the advantage over directed gene approaches of looking at the expression of hundreds to thousands of genes simultaneously. As a result, it is possible to define sets of genes related to structural, physiological or metabolic functions that could underlay the differences observed between fish ecotypes. For example, using GRASP (Genomic Research on Atlantic Salmon Project) chips, differences in genes involved in energy, lipid metabolism, muscle contractility and growth were differentially regulated between sympatric whitefish ecotypes that exhibited differences in habitat and life history strategies (Derome & Bernatchez 2006; Derome *et al.* 2006; St-Cyr *et al.* 2008). We have proposed that, given the advances in deep sequencing platforms and technologies, deep sequencing should replace the use of hybridization-based approaches (i.e. microarrays) for quantitative comparative transcriptomics, particularly for nonmodel species (Goetz & MacKenzie 2008). There are several advantages of transcriptome sequencing (RNA-seq) compared to hybridization-based approaches. With RNA-seq, there are no preconceived ideas of which genes are important. Hybridization arrays are only as good as the genes they contain, while RNA-seq could theoretically provide an entire transcriptomic profile and thereby enable transcript quantification and gene discovery to occur simultaneously. This is particularly important for nonmodel organisms for which sequenced genomes do not exist. In addition, when studying organisms with duplicated genomes and/or strains with genetic dissimilarity (as in the current study), differences in hybridization efficiency can impact results.

While next-generation sequencing platforms have the capacity to produce millions of reads per sequencing run, some platforms such as the Illumina-Solexa, ABI SOLiD and HeliScope, produce sequences with short read lengths. This makes transcript identification by sequence alignment difficult for nonmodel organisms or species without genomic sequences for comparison. Effective bioinformatics such as *de novo* assembly and expression profiling also present challenges with short

read lengths. In contrast, the Roche 454 Genome Sequencer (GS)-FLX produces relatively long read lengths (>200 bp) compared to other next-generation sequencing approaches, benefiting assembly and annotation. In addition, investigators have demonstrated that 454 sequencing technology has the potential to be effectively used for gene expression profiling (Torres *et al.* 2008). Thus in the current study, a second objective was to use the Roche GS-FLX 454 sequencing technology to compare the hepatic transcriptomes between lean and siscowet lake trout that were raised under identical environmental conditions and were the same fish used to analyse growth, morphometrics and lipid content. We hypothesized that, given the differences in lipid content in the muscle, we would observe differences in the transcript levels of hepatic genes involved with lipid metabolism, synthesis and binding between lean and siscowet lake trout. Further, given the depth of sequencing provided by pyrosequencing, we would be able to use this technology to quantitatively compare differences in gene read frequencies between the siscowet and lean lake trout livers.

Methods

Animals, husbandry and rearing

On 26 October 2006, siscowet lake trout were sampled at a bottom depth of 109 m using multifilament bottom-set gill nets at an offshore site in Lake Superior northeast of Marquette, MI. Siscowet were chosen based on their morphometry including the shape and relative size of the head, size of the fins and location and size of the eyes (Moore & Bronte 2001), and the fact that they were collected at depths greater than 100 m (Bronte *et al.* 2003). Lake trout forms designated as siscowets, leans and humpers have been shown to be genetically separated based on microsatellite analysis (Page *et al.* 2004). Fish averaged 2184 g in weight and a number of females sampled had ovulated eggs in the body cavity and males had easily expressible milt. Eggs of 12 females were fertilized in separate batches with milt from two to three males using the wet method. After washing, eggs were placed in Teflon containers and were transported to the Great Lakes WATER Institute (GLWI) on ice then incubated in a Heath vertical incubator with flow-through water at 10 °C in the dark. All eggs hatched by 7 January 2007.

On 21 December 2006, newly hatched lean lake trout arrived at the GLWI from the Les Voigt Fish Hatchery (Wisconsin DNR). These fry originated from spawning 24 female and 24 male wild lean lake trout caught 13 October 2006 on the Gull Island Shoal in western Lake Superior. Thus, the lean lake trout were approximately 2 weeks older than the siscowet. Lean lake trout fry were incubated in separate trays (with lids) within the same Heath vertical incubator that held the siscowet eggs. Lean lake trout swim-up fry were moved from the incubator to a 4 ft. diameter tank on 18 January 2007 and siscowet fry were moved to a separate 4 ft. diameter tank on 1 February 2007. These tanks were supplied with flow-through water at 10 °C plumbed from the same header tank. Both siscowet and lean lake trout fry were feed habituated to Rangen Trout Starter diet (Rangen Inc.) using automatic vibratory feeders set for 21 feedings/day and were transitioned through Rangen trout crumble and pellets as they grew.

At 6 months posthatch, 500 lean and 500 siscowet lake trout were obtained randomly from the stock tanks and placed into two 1000 litre tanks (250 lean lake trout and 250 siscowet lake trout comingled/tank). Lean lake trout were identified with an adipose fin clip. At 2 years of age, approximately 80 lean and 80 siscowet lake trout were moved from each tank into a third 1000 litre tank to decrease overall fish densities. All tanks initially received 4% of their body weight (Rangen Trout Production pellets) per day that was distributed by three automatic vibratory feeders per tank over 15 feedings/day and was gradually reduced to 0.75%, coinciding with decreased feed intake as fish got older. Gut analyses on a subset of fish terminally sampled at 1 and 2 years of age, indicated that all fish were feeding. Fish were sampled for digital photography, lipid analysis and tissues for genomic assays at 1 and 2 years of age as described below.

Growth

Measures of length (0.25 cm) and weight (0.1 g) were taken on a subsample of 25 fish of each lake trout form per tank when the fish were comingled on 18 June 2007 and repeated every other week from June to August (2007), and then once monthly thereafter. For these assays, sampled fish were anesthetized with 100 mg/L MS-222 (Argent Labs), measured, and returned to the tank for recovery. On 16 September 2008 and 5 January 2009, lengths and weights were conducted on all the fish in each tank. Trends in growth (weight and length) and condition factor were assessed from the periodic sampling but were only analysed for statistical significance (two-sample *t*-test) at the two assays where all fish were measured. Condition factor (K_{TL}) was calculated using:

$$K_{TL} = 100000W/L^3,$$

where W is the weight in grams and L is total length in millimetres (Carlander 1950).

Truss analysis

The truss protocol (Strauss & Brookstein 1982; Brookstein *et al.* 1985) was used to describe differences in the shape of specific regions of each fish. Trusses are vertical, horizontal and oblique distances measured between preselected anatomical landmarks (Fig. 2), which are points identified on the basis of local morphological features and chosen to divide the body into functional units (Brookstein *et al.* 1985). For truss analysis, digital photographs of the whole bodies (left lateral side) of 10 and 15 individuals per lean and siscowet lake trout were taken at 1 and 2 years posthatch, respectively. All landmark and truss measures used were taken from prior studies directed at the delineation of Lake Superior lean and siscowet lake trout (Moore & Bronte 2001; Bronte & Moore 2007). Truss measurements were collected from the photographs using ImageJ (1.32j) and were analysed using the multivariate analysis of variance (MANOVA) function in Minitab (15.1.30.0). If the MANOVA indicated significant differences between lean and siscowet lake trout, univariate analysis of variance was completed to identify the truss elements that were significantly different ($P \leq 0.05$).

Lipid analysis

In year 1 of the study, six lean and six siscowet lake trout were sampled as whole fillets minus heads and viscera. Sampled fish were all within 5% of the mean length for each lake trout form (lean = 17.00 ± 0.25 cm and 40.3 ± 2.2 g; siscowet = 16.75 ± 0.25 cm and 43.8 ± 2.3 g). Skin-on fillets were oven dried for 2 days at 104 °C prior to grinding and lipid analysis. In year 2, a 1 in. steak cross-section of muscle was sampled from 20 lean and 20 siscowet lake trout just posterior of the head and pectoral fins. Sampled fish were all within 5% of the mean length for each lake trout form (lean = 39.00 ± 1.00 cm and 608.0 ± 20.9 g; siscowet = 39.25 ± 0.75 cm and 630.4 ± 20.6 g). The steak cross-section was similarly minced and oven dried for 2 days at 104 °C prior to grinding. The high lipid content of the samples in both years necessitated the use of dry ice pellets while grinding to ensure a homogeneous sample and to prevent grinding to a paste. After grinding, samples were dried at 104 °C overnight to remove excess carbon dioxide and moisture that accumulated while grinding.

Lipids were extracted from dried samples via the Soxhlet method (AOCS Method Ba 3-38) with a Büchi 810 Fat Extraction Apparatus. Sample size was approximately 2 g and extractions lasted 2 h at a drip rate of over 200 drips/min. Extracted lipids were dried to constant weight at 104 °C and fat content was determined on a gravimetric basis and expressed as percent by wet or dry weight of sample in the results. For statistical analysis, data were arcsin transformed and analysed by analysis of variance (ANOVA) followed by Tukey's post-hoc test. Pairwise comparisons with *P*-values of ≤0.05 were considered significantly different.

Fatmeter

At 27 months, 19 siscowet and 19 lean lake trout within 5% mean length for each lake trout form (siscowet = 984.4 ± 59.0 g and 44.75 ± 1.00 cm; lean = 856.4 ± 52.0 g and 43.75 ± 1.25 cm) were analysed for lipid content by microwave. Fish were briefly anaesthetized by immersion in MS222 (100 mg/L), weighed and measured for length, and analysed for lipid content. Each fish was analysed for lipid content on five defined areas on the right side of the body (Fig. 4) using a noninvasive microwave sensor device (Distell Model 692 Fish Fatmeter, Distell Inc.). The device emits a low-powered wave (2 GHz, 2000 MHz, power 2 mW) that interacts with water within the somatic tissues and uses the inverse relationship between water and lipid to estimate the lipid concentration in tissue (as a percent) from each of the specified locations. Each reading took less than 30 s to register. The locations on the fish were chosen based on the manufacturer's recommendation and prior research suggested that the combination of these readings provided a reasonable representation of total lipid content in the fish (Crossin & Hinch 2005). Previous research also demonstrated no significant differences between measurements taken at the specified location on either side of the fish (C. A. Murphy, unpublished). The manufacturer provides species-specific calibrations that are programmed into the fatmeter, however, the 'research' setting was used since there were no programmed calibrations for lake trout. In previous research, calibration using Soxhlet extractor showed a strong linear log transformed relationship between device output and lipid content ($R^2 = 0.78$; C. A. Murphy, unpublished) and that different calibrations for siscowet and lean lake trout were not required (M. Zimmerman, personal communication).

Since we were primarily interested in differences in the relative lipid content between lean and siscowet lake trout and in whether lipid accumulated in different areas of the body, the device readings were not transformed to represent absolute lipid concentrations. In addition, normalization for differences in body size that might occur in wild sampled fish were not necessary since all fish sampled were within 5% of the mean length. Fatmeter readings (% lipid) from each of the locations were first arcsin transformed and compared between the two lake trout forms using a MANOVA function in Minitab (15.1.30.0). If the MANOVA indicated

significant differences between the lake trout forms, univariate analysis of variance was completed to identify the sample locations that were significantly different ($P \leq 0.05$).

RNA extraction, cDNA preparation and pyrosequencing

For samples taken in years 1 and 2, liver sections were dissected, snap frozen and stored at -80 °C until processed. Following thawing, livers were extracted in Tri Reagent (Molecular Research Center, Inc.) according to the manufacturer's protocol (Chomcynski & Sacchi 1987; Chomcynski 1993). The RNA was treated with DNase I and cleaned using the RNeasy MinElute Cleanup kit (Qiagen). Pyrosequencing was only conducted on the RNA extracted from liver samples taken at year 1. For cDNA synthesis, equal quantities of total RNA were combined between 10 liver samples per lake trout form (lean and siscowet). Because large quantities (10 μg/morphotype) of cDNA are required for 454 sequencing, the SMART PCR cDNA synthesis kit (Clontech) was used according to the manufacturer's instructions. This kit initially produces single-stranded cDNA from RNA by reverse transcription followed by PCR amplification to produce double-stranded cDNA. Subsamples of cDNA from lean and siscowet lake trout were examined using agarose gels to ensure that the same relative amounts of cDNA were being produced. Pyrosequencing was conducted on a Roche GS-FLX (454 Life Sciences) at the Interdisciplinary Center for Biotechnology Research genomics core at the University of Florida (Gainesville, FL). Libraries were constructed for leans and siscowets using unique tags for each library; Ligation Multiplex Identifiers (MIDs). The two libraries were combined and titred by sequencing approximately 16 000 sequences. The titres on the initial sequencing run indicated a ratio of 1:5 sequences (MID1 (siscowet):MID2 (lean)). In the final full-scale sequencing run the ratio of sequences between siscowet and lean libraries was 1:1.75 and this value was used to adjust the gene frequencies when comparing sequence reads between siscowet and lean lake trout (Tables 1–3). All of the original Roche 454 GS-FLX sequences for this study have been submitted to the NCBI Short Read Archive for public access (study SRP001186).

Sequence analysis

Sequences from the two libraries were initially separated by respective MIDs at the genome core facility. At the GLWI, the sequences were further processed to remove linker primer sequences derived from the SMART cDNA synthesis. In addition, poly A/T tails were also trimmed from the sequences when they occurred at the end or the beginning of sequences.

For the comparative analysis of gene expression between lean and siscowet lake trout, two bioinformatic approaches were employed, one using BLASTN alignments and one using assembled contigs. At the beginning of this investigation, programs that could effectively assemble the total number of gene reads produced from the GS-FLX sequencing of unnormalized cDNA were not available. Thus, we developed an approach in which the individual gene reads were compared to each other using BLASTN and the top 500 hits were recorded/sequence. Sequences were then grouped into gene clusters based on their BLAST score and the relative number of sequences from both MID libraries in each group was calculated. Sequences within each gene cluster were then assembled using CAP3 (Huang & Madan 1999) for subsequent annotation and qPCR primer design (see below). For this analysis, only sequences ≥100 bp after trimming were analysed. This amounted to 91 810 sequences for siscowet lake trout and 160 892 sequences for lean lake trout. Later in our study, commercial platforms were released that contained programs that could assemble the very large sequence sets produced from next-generation sequencing of unnormalized cDNA. Thus, in a second approach, *de novo* assembly of sequencing reads was performed, followed by comparative expression analyses (Genomic Workbench, CLCBio). Initially, all sequences were trimmed based on quality scores of 0.05 (Phred; Ewing & Green 1998; Ewing *et al.* 1998) and the number of ambiguous nucleotides (>2 on ends). Sequences smaller than 30 bp were removed leaving 153 046 sequences for the siscowet library and 272 775 sequences for the lean library. Assembly resulted in 2276 contiguous sequences which were then used as a reference for comparing expression between the siscowet and lean MID libraries in two ways: (i) ratios of gene reads per MID library, and (ii) relative expression values measured in RPKM (reads per kilobase of exon model per million mapped reads) (Mortazavi *et al.* 2008) between MID libraries.

For all bioinformatic approaches, assembled contigs were aligned by BLAST to the NCBI nr protein, nucleotide and dbEST ('all others') databases for annotation, and sequences and detailed BLAST results are provided in the Supplemental Data (Contig Sequences and Annotation).

Quantitative reverse transcription-polymerase chain reaction

Quantitative reverse transcription-polymerase chain reaction (qPCR) was carried out to assess the results obtained from the various bioinformatic approaches used to analyse the RNA-seq data. Primers for qPCR

Table 1 Gene frequency analysis from contiguous assemblies obtained from BLASTN alignment analysis and qPCR corroboration

Leans over siscowets	Gene frequency analysis						QPCR analysis			
Gene annotation	Contig no.	Total # seq.	Lean seq.	Siscowet seq.	Lean (L)/ siscowets (S)	Adjusted L over S	Lean mean	Siscowet mean	Difference L–S	P-value
C1q-like adipose specific protein	9821	135	129	6	21.5	12.3	23.31	0.81	22.50	0.039
ovary-specific C1q-like factor	26546	19	18	1	18	10.3	10.01	6.22	3.78	0.001
similar to complement component C3	16894	19	18	1	18	10.3	13.25	5.65	7.60	0.014
proteasome subunit alpha type-2	14019	19	18	1	18	10.3	2.89	2.14	0.75	0.008
RING-box protein 1	9356	19	18	1	18	10.3	11.92	9.72	2.20	0.039
NADH dehydrogenase 1	24752	15	14	1	14	8.0				
Acyl-CoA desaturase	22337	14	13	1	13	7.4	875.35	642.06	233.28	0.060
Secreted immunoglobulin domain 4	22567	14	13	1	13	7.4	1038.02	963.70	74.32	0.465
Unknown but blasts against EST	26150	13	12	1	12	6.9	8.55	8.59	−0.05	0.982
Unknown (no blast hits)	24348	13	12	1	12	6.9				
Solute carrier family 27	27857	12	11	1	11	6.3	32.52	31.17	1.35	0.783
Reticulon 4	26561	12	11	1	11	6.3	38.24	28.87	9.36	0.034
Apolipoprotein A-IV precursor	25572	12	11	1	11	6.3	10.62	4.76	5.85	0.136
Catechol-O-methyltransferase	24744	12	11	1	11	6.3	0.00	0.00	0.00	0.520
Unknown (blasts to ESTs)	15409	12	11	1	11	6.3	4.39	2.33	2.06	0.001
Mitochondrial ribosomal protein S33	11265	12	11	1	11	6.3	8.35	5.63	2.73	0.030
Cytochrome c oxidase polypeptide VIII- heart	8226	12	11	1	11	6.3	40.41	41.08	−0.67	0.890
Hypothetical S. salar protein	7115	12	11	1	11	6.3	10.44	9.14	1.30	0.150
Diacylglycerol O-acyltransferase 2	5276	12	11	1	11	6.3	21.79	25.66	−3.87	0.358
Apolipoprotein B	25963	11	10	1	10	5.7	47.26	35.19	12.06	0.025
glycine C-acetyltransferase	20462	11	10	1	10	5.7	8.41	7.81	0.60	0.520

Siscowets over leans										
Gene annotation	Contig no.	Total # seq.	Siscowet seq.	Lean seq.	Siscowets (S)/ leans (L)	Adjusted S over L	Siscowet mean	Lean mean	Difference S–L	P-value
Stannin	4651	18	17	1	17	29.8	0.18	0.20	−0.02	0.295
Unknown (no blast hits)	10018-1	11	10	1	10	17.5	23.87	26.05	−2.18	0.050
Unknown (no blast hits)	10018-2									
Guanine nucleotide binding proteinβ2	2046	21	19	2	9.5	16.6	68.44	61.13	7.31	0.235
Arginine rich protein (ARMET)	3316	12	10	2	5	8.8	5.40	8.28	−2.87	0.018
Unknown (blasts to ESTs)	1325	11	9	2	4.5	7.9	1.06	1.35	−0.28	0.189
Trafficking protein particle complexsubunit 2-like protein	3108	11	9	2	4.5	7.9	2.49	2.61	−0.12	0.753
FK506-binding protein 5	4479	11	9	2	4.5	7.9	0.17	0.31	−0.14	0.005
Similar to Fibrocystin-L precursor	766	11	9	2	4.5	7.9	0.04	0.04	−0.01	0.782
Aginase type Ib	2730	10	8	2	4	7.0	3.32	3.06	0.26	0.518
Hypothetical protein	983	14	11	3	3.67	6.4				
Replication protein A	5080	14	11	3	3.67	6.4	0.33	0.33	0.01	0.890
Perforin	549	22	17	5	3.4	6.0	43.57	24.40	19.17	0.653
ATP synthase lipid-binding protein	374	13	10	3	3.33	5.8	29.19	37.49	−8.30	0.053
Solute carrier family 25 alpha	1809-1	13	10	3	3.33	5.8	5.37	4.85	0.52	0.245
Zonadhesin-like gene	2198	13	10	3	3.33	5.8	7.18	9.37	−2.19	0.192
Alpha-tectorin-like protein	2247	13	10	3	3.33	5.8				
Myosin 1	3907	13	10	3	3.33	5.8	25.43	29.41	−3.98	0.187
Cytochrome P450	2080	17	13	4	3.25	5.7	8.96	7.17	1.79	0.165

Sequences for contigs are provided in Supplemental data with detailed top BLAST information for gene annotation. 'Adjusted' values represent either (S over L) × 1.75 or (L over S)/1.75 to account for differences in the overall number of sequences in lean and siscowet Multiplex Identifiers libraries. qPCR (quantitative reverse transcription–polymerase chain reaction) mean values are normalized (to actin) expression values.

(Supplemental Data: qPCR primers) were designed to contiguous assemblies (Supplemental Data: Contig Sequences) and qPCR was performed on individual RNA samples from the same 10 lean and 10 siscowet lake trout that were pooled originally to generate cDNA for 454 sequencing. First-strand cDNA synthesis was performed on total RNA with Improm II (Promega) reverse transcriptase using 0.5 μg RNA, 0.25 μg oligo dT primer and 10 mM dNTP mix. Reactions were carried out for 1 h at 37 °C. All qPCR reactions were

Table 2 Gene frequency analysis from contiguous assemblies produced by Genomics Workbench (CLCBIO) software and qPCR (quantitative reverse transcription–polymerase chain reaction) corroboration

Leans over siscowets	Gene frequency analysis					qPCR analysis			
Gene annotation	Contig no.	Total # seq.	Lean seq.	Siscowet seq.	Lean (L)/ siscowets (S)	Adjusted L over S	Lean mean	Siscowet mean	Difference L–S	P-value
C1q-like adipose specific protein	25	126	120	6	20.0	11.4	23.31	0.81	22.50	0.039
Proteasome subunit alpha type-2	1404	19	18	1	18.0	10.3	2.89	2.14	0.75	0.008
Secreted immunoglobulin domain 4	895	18	17	1	17.0	9.7	91.77	32.09	59.69	0.236
RING-box protein 1	1867	17	16	1	16.0	9.1	11.92	9.72	2.20	0.039
Unknown but blasts to ESTs	1142	14	14	0	14.0	8.0	1.21	1.32	−0.11	0.501
Acyl-CoA desaturase	2062	28	26	2	13.0	7.4	875.35	642.06	233.28	0.060
N-acetylglucosamine-6-sulfatase precursor	1416	13	13	0	13.0	7.4	1.15	1.07	0.08	0.535
Unknown (no blast hits)	1575	13	13	0	13.0	7.4	0.28	0.17	0.10	0.032
Microsomal glutathione S-transferase 3	1371	62	57	5	11.4	6.5	26.02	14.35	11.67	0.001
C1 inhibitor	1202	32	29	3	9.7	5.5	501.86	453.56	48.30	0.182
Pyridoxine-5-phosphate oxidase	1157	17	15	2	7.5	4.3	0.28	0.17	0.10	0.032
Unknown (no blast hits)	1664	41	36	5	7.2	4.1	1.50	2.97	−1.47	0.374
Fatty acid binding protein, heat	458	31	27	4	6.8	3.9	41.79	42.23	−0.43	0.958
Complement C1q-like protein 2	1225	306	265	41	6.5	3.7	613.16	66.97	546.19	0.011
Peroxisome proliferator activated receptor	637	321	269	52	5.2	3.0	26.81	19.82	6.99	0.010

Siscowets over leans									
Gene annotation	Contig no.	Total # seq.	Siscowet seq.	Lean seq.	Siscowets (S)/ leans (L)	Adjusted S over L	Siscowet mean	Lean mean	Difference S–L	P-value
CDV3 homolog	1396	28	27	1	27.0	47.3	1.18	1.33	−0.15	0.341
Similar to FBP32	732	24	23	1	23.0	40.3	54.97	53.32	1.65	0.818
Stannin	1147	17	16	1	16.0	28.0	1.08	0.96	0.12	0.401
Ribosomal protein L17	1754	17	16	1	16.0	28.0	45.18	48.97	−3.78	0.613
Unknown (no blast hits)	493	14	14	0	14.0	24.5	0.03	0.04	−0.01	0.042
Unknown (blasts to ESTs)	1013	11	11	0	11.0	19.3	CT values to low for calculation			
Src family associated phosphoprotein 1	1180	11	10	1	10.0	17.5	0.39	0.54	−0.15	0.034
Unknown (blasts to ESTs)	1695	9	8	1	8.0	14.0	0.05	0.05	0.00	0.591
Anionic trypsin-1 precursor	1242	9	8	1	8.0	14.0	5.36	3.49	1.87	0.434
Microsatellite Alu16 sequence	340	9	8	1	8.0	14.0	0.09	0.10	−0.01	0.750
Unknown	307	9	8	1	8.0	14.0	23.87	26.05	−2.18	0.050
Similar to FBP32	169	35	31	4	7.8	13.6	10.41	9.20	1.21	0.201
Unknown (blasts to ESTs)	1093	7	7	0	7.0	12.3	7.60	7.82	−0.23	0.863
Unknown (blasts to ESTs)	908	12	10	2	5.0	8.8	2.63	2.77	−0.15	0.784
FK506-binding protein 5	1390	12	10	2	5.0	8.8	0.17	0.31	−0.14	0.005

Sequences for contigs are provided in Supplemental data with detailed top BLAST information for gene annotation. 'Adjusted' values represent either (S over L) × 1.75 or (L over S)/1.75 to account for differences in the overall number of sequences in lean and siscowet Multiplex Identifiers libraries. qPCR means are normalized (to actin) expression values.

Table 3 Analysis of contiguous assemblies having the highest overall expression values when analyzed by Genomics Workbench (CLCBio) software

Lean over siscowet

Gene annotation	Gene frequency analysis						Gene expression analysis			qPCR analysis			
	Contig no.	Total # seq.	Lean seq.	Siscowet seq.	Lean (L)/siscowet (S)	Adjusted L over S	Lean expression	Siscowet expression	Expression difference	Lean mean	Siscowet mean	Difference L–S	P-value
28S Ribosomal RNA	1367	847	669	178	3.8	2.1	14 450	6752	7698	110.59	82.81	27.77	0.041
Apolipoprotein AII	1472	1830	1352	478	2.8	1.6	16 286	10 112	6174	103.82	91.47	12.35	0.179
Cysteine sulfinic acid decarboxylase	8	11576	7772	3804	2.0	1.2	35 840	30 805	5035	9.71	9.12	0.59	0.751
Mitochondrial genome blast	1603	10671	7121	3550	2.0	1.1	29 572	25 890	3682	2.85	1.80	1.05	0.005
B-microseminoprotein precursor	2068	786	559	227	2.5	1.4	11 892	8480	3412	3.06	2.17	0.89	0.027
Coiled-coil transcriptional coactivator b	1423	1536	1104	432	2.6	1.5	10 685	7342	3343	12.78	11.19	1.59	0.207
Complement C1q-like protein 2	1225	306	265	41	6.5	3.7	3968	1078	2890	613.16	66.97	546.19	0.011
Mitochondrial genome blast	115	1915	1321	594	2.2	1.3	12 862	10 156	2706	17.87	16.17	1.70	0.474
Liver SSH library from warm, cold and hypoxic stress	1506	518	394	124	3.2	1.8	5220	2885	2335	333.12	176.35	156.77	0.035
Serum albumin 2	485	6055	4025	2030	2.0	1.1	19 782	17 521	2261				
Liver-type fatty acid binding protein	662	1684	1118	566	2.0	1.1	20 331	18 067	2264	76.56	70.90	5.66	0.475
Cytochrome C oxidase subunit II	2001	1246	841	405	2.1	1.2	12 609	10 664	1945				
Apolipoprotein A-I-1	1190	1118	776	342	2.3	1.3	7940	6145	1795	105.41	96.58	8.84	0.402
Unknown but blasts strongly to ESTs	2078	513	354	159	2.2	1.3	7780	6137	1643				
Catechol-O-methyltransferase domain-containing protein 1	1214	245	203	42	4.8	2.8	2466	896	1570	30.49	12.56	17.93	0.001
Alpha 1-antiproteinase-like protein	133	656	485	171	2.8	1.6	3816	2363	1453	15.41	15.65	-0.24	0.887
Peroxisome proliferator activated receptor gamma	637	321	269	52	5.2	3.0	2165	735	1430	26.81	19.82	6.99	0.010
60s Ribosomal protein L36	1855	130	100	30	3.3	1.9	2891	1523	1368	1.28	1.15	0.13	0.4

Siscowet over lean

Gene annotation	Contig no.	Total # seq.	Siscowet seq.	Lean seq.	Siscowet (S)/lean (L)	Adjusted S over L	Siscowet expression	Lean expression	Expression difference	Siscowet mean	Lean mean	Difference S–L	P-value
Fucolectin-4 precursor	1973	5424	2354	3070	0.8	1.3	39 701	29 484	10 217	29.04	23.11	5.93	0.317
Saposin-related isoform A/antimicrobial peptide NK-lysin	1441	1385	765	620	1.2	2.2	17 654	8147	9507	74.78	101.76	-26.98	0.070
Saposin-related, isoform A NK-lysin type 3	1871	1359	768	591	1.3	2.3	13 849	6069	7780	2734.52	2373.41	361.10	0.635
Type-4 ice-structuring protein LS-12 precursor	1942	827	473	354	1.3	2.3	13 341	5686	7655	1013.74	645.76	367.98	0.012
RNAse2/angiogenin	1450	1529	730	799	0.9	1.6	17 713	11 040	6673	6.17	7.55	-1.38	0.334
Precerebellin	2013	1667	747	920	0.8	1.4	19 385	13 595	5790	4.35	5.45	-1.10	0.074
Serum albumin 2	2049	3894	1701	2193	0.8	1.4	20 971	15 396	5575				
Hemopexin-like protein	1733	2494	1180	1314	0.9	1.6	15 025	9528	5497	20.82	23.95	-3.13	0.122

Table 3 *Continued*

Siscowet over lean	Gene frequency analysis						Gene expression analysis			qPCR analysis			
Gene annotation	Contig no.	Total # seq.	Siscowet seq.	Lean seq.	Siscowet (S)/lean (L)	Adjusted S over L	Siscowet expression	Lean expression	Expression difference	Siscowet mean	Lean mean	Difference S–L	P-value
Saposin-related antimicrobial peptide NK-lysin	1856	1157	562	595	0.9	1.7	10 741	6475	4266	820.47	833.20	−12.74	0.957
Acyl CoA binding protein	1466	505	263	242	1.1	1.9	6358	3332	3026	1341.86	1031.22	310.65	0.052
Similar to neurotoxin/C59/Ly6	1148	425	238	187	1.3	2.2	5379	2407	2972	313.01	249.88	63.13	0.085
Secreted phosphoprotein 24	1809–2	1271	549	722	0.8	1.3	11 764	8810	2954	9.78	9.03	0.75	0.542
Fucolectin-4 precursor	1253	673	307	366	0.8	1.5	7228	4907	2321	2384.05	2376.82	7.23	0.988
CDV3 homolog	1396	27	27	0	27.0	47.3	2147	0	2147	1.18	1.33	−0.15	0.341
SE-chephalotoxin	2102	1006	490	516	0.9	1.7	4280	2566	1714	4.61	5.75	−1.13	0.345
Haptoglobin	71	339	192	147	1.3	2.3	2970	1295	1675	5.24	3.04	2.20	0.044
Complement factor H	1822	93	59	34	1.7	3.0	2018	662	1356	1.76	1.66	0.10	0.763
Toxin 1	1795	330	140	190	0.7	1.3	5739	4435	1304				

For gene frequency analysis, adjusted values represent either (S over L) × 1.75 or (L over S)/1.75 to account for differences in the overall number of sequences in lean and siscowet Multiplex Identifiers libraries. qPCR (quantitative reverse transcription–polymerase chain reaction) means are normalized (to ac tin) expression values. Sequences for contigs are provided in Supplemental data with detailed top BLAST information for gene annotation.

created as master mixes and individual reactions contained the following: 2.5 μL of a 1:10 dilution of cDNA, 5 pM each of forward and reverse gene primers (Supplemental Data: qPCR primers), and 12.5 μL Power SYBR Green PCR Master Mix (Applied Biosystems). Cycling and fluorescence measurements were carried out in a Stratagene Mx 3000P System (Stratagene) with the following cycling parameters: 1 cycle of 95 °C for 10 min; 40 cycles of 95 °C for 15 s and 58 °C for 1 min. Fluorescence readings were taken at the end of each cycle. Immediately after cycling, a melting curve analysis was run. Amplification products from qPCR primers were analysed initially on agarose gels to ensure the presence of single bands of the correct size, and quality control for qPCR included the analysis of no template controls for the absence of primer dimers, and dissociation curves for the presence of sharp single peaks.

Raw data were processed with Real-time PCR Miner (Zhao & Fernald 2005). Quantification was performed by calculating the relative mRNA concentration (R_0) for each gene per individual sample. Briefly, this was calculated using the following equation: $R_0 = 1/(1 + E)^{Ct}$, where E is the gene efficiency calculated as the average of all individual sample efficiencies across all reactions for a given gene per qPCR plate, and Ct is the cycle number at threshold (Zhao & Fernald 2005). The R_0 for each gene was normalized to a control (actin—Supplemental Data: qPCR primers). Average normalized values were calculated for lean and siscowet lake trout per gene and analysed by ANOVA. P-values of ≤0.05 were considered significantly different, though P-values for all reactions are provided in the results and were sometimes used more broadly to interpret qPCR corroboration.

Results

Growth

When lean and siscowet lake trout began communal rearing (6 months posthatch), lean lake trout were 6.9 ± 1.3 g and 9.75 ± 0.50 cm, and siscowets were 4.6 ± 1.0 g and 8.50 ± 0.50 cm. Lean lake trout were 2 weeks older and initially were slightly larger then siscowets. Even so, the siscowets outgrew the leans (Fig. 1). By 2–2.5 years old, when all fish were weighed, the siscowets were significantly heavier (859.9 g $t = -6.83$, d.f. = 810; $P < 0.001$) and longer (42.50 cm, $t = -6.64$, d.f. = 773; $P < 0.001$) than leans (746.9 g and 41.1 cm), however the difference in length was minor compared to the difference in weight (Fig. 1). As a result, siscowet ($K_{TL} = 1.10$) also had a significantly ($t = -9.36$, d.f. = 785, $P < 0.001$) higher condition factor than lean lake trout ($K_{TL} = 1.05$).
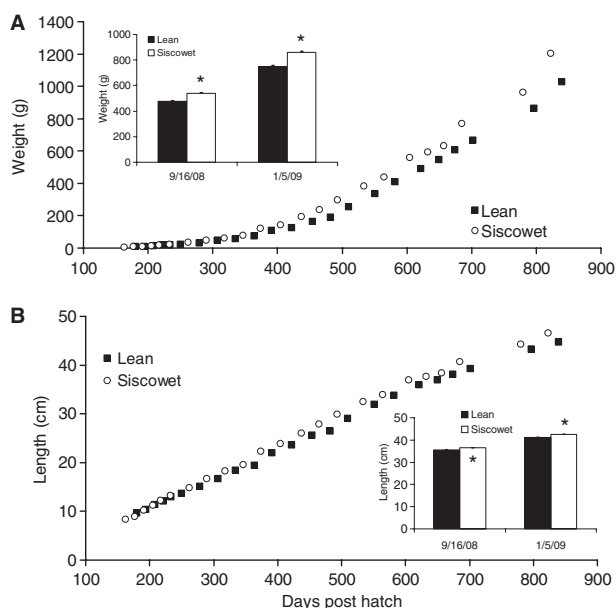
Fig. 1 Mean weights (A) and lengths (B) of laboratory-reared lean and siscowet lake trout calculated from assays performed semimonthly (first 3 months) and monthly (>3 months) for 2.5 years. Each point represents the mean for 25 individuals/lean or siscowet lake trout/tank. Inset figures indicate the mean weight (g) ± SD and length (cm) ± SD assayed at two time points for all individuals/lean or siscowet lake trout (16 September 2008: Lean $N = 412$; Siscowet $N = 427$; 5 January 2009: Lean $N = 395$; Siscowet $N = 393$). Asterisks indicates significant difference between lean and siscowet lake trout at $P \leq 0.05$.

Morphometry

At 1 year, three truss elements (1, 3, 7) were significantly different and by 2 years, this increased to six elements (1, 2, 3, 4, 5, 9), mostly associated with head, body depth and caudal peduncle measures (Fig. 2).

Lipids

Lipids were analysed in lean and siscowet lake trout by a chemical method and by a microwave device. While the chemical method is a precise measure of the lipids in the muscle, the microwave device allowed us to quickly compare relative lipid levels at a number of different areas of the body. At both sampling times, siscowet lake trout had a significantly higher percentage of lipid compared with lean lake trout on both a dry or wet weight basis (Fig. 3). In addition, there was a significant increase in lipids from year 1 to year 2 for both lean and siscowet lake trout on a wet or dry weight basis. At all five locations sampled with the fatmeter, there was a significantly greater reading in siscowet than in lean lake trout, indicating a higher percentage of fat at those locations (Fig. 4B). In many cases, there were significant differences between the fatmeter read-

ings at different body locations within each lake trout form, however this was observed more in siscowet lake trout (Fig. 4C). For example, readings at locations 1 and 2 in siscowet lake trout were very different ($P < 0.0001$) from location 3, whereas in lean lake trout, readings at location 1 were not different from location 3 and readings at location 2 were just significantly ($P < 0.045$) different from readings at location 3.

Comparative transcriptomics

The analysis of sequences using BLASTN alignments identified 21 gene clusters containing sequences having a higher (≥ 5.7) adjusted gene frequency in lean vs. siscowet libraries, and 19 gene clusters containing sequences with higher (≥ 5.7) adjusted gene frequencies in siscowet compared with lean lake trout (Table 1). qPCR results corroborated 10 (one at $P = 0.06$) of the lean over siscowet gene frequency comparisons, and with a few exceptions, even the ones not corroborated by qPCR were in the correct direction (e.g. lean qPCR mean > siscowet qPCR mean—Table 1) as expected from the gene frequency analysis. However, qPCR did not corroborate the gene frequency results obtained from the BLASTN groupings in which siscowet lake trout had higher gene frequencies than lean lake trout. In fact, four gene expression profiles were significantly different in the opposite direction (e.g. leans > siscowets) to what was predicted from the gene frequency results (Table 1). Regardless, the differences in expression measured by qPCR were not very large between the lake trout forms for those genes with higher frequencies in siscowet lake trout. In contrast, the mean differences observed in expression between some of the genes with higher frequencies in lean vs. siscowet lake trout were large (e.g. C1q adipose-specific protein).

Gene frequency analysis derived from aligning individual reads to 2276 assembled contiguous sequences (Genomic Workbench, CLCBio) produced similar results to the BLASTN gene clustering. In this analysis, the 15 assembled contigs with the greatest gene frequency differences for each MID library comparison (lean over siscowet and siscowet over lean) were analysed using qPCR (Table 2). Of the genes exhibiting higher frequencies in lean vs. siscowet lake trout, nine of the comparisons were corroborated by qPCR with many of them having large fold differences in expression levels (Table 2). Further, several of these qPCR-corroborated genes were the same as those observed in the BLASTN-based analysis including, C1q-like adipose-specific protein, proteasome subunit alpha type-2, RING-box protein 1 and acyl-CoA desaturase. In addition, the frequency differences for these genes were in the same order for both analyses (Tables 1 and 2). Again, however, qPCR results did not agree with results for genes that exhibited
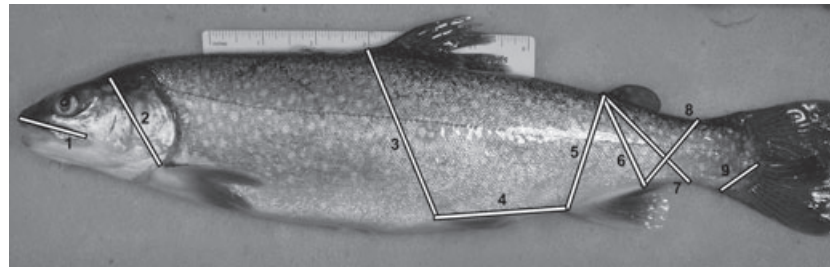
| Truss | Year 1 | | | Year 2 | | |
	Lean average (cm), SD	Siscowet average (cm), SD	P-value	Lean average (cm), SD	Siscowet average (cm), SD	P-value
Truss 1	1.86, 0.15	1.67, 0.10	0.003 ∗	3.60, 0.19	3.43, 0.20	0.033 ∗
Truss 2	2.52, 0.10	2.55, 0.08	0.449	4.98, 0.20	5.17, 0.16	0.008 ∗
Truss 3	3.84, 0.17	4.08, 0.14	0.006 ∗	8.19, 0.13	8.54, 0.22	0.002 ∗
Truss 4	3.29, 0.16	3.23, 0.19	0.380	6.81, 0.30	6.36, 0.32	<0.001 ∗
Truss 5	2.52, 0.09	2.60, 0.11	0.120	5.37, 0.09	5.57, 0.15	0.005 ∗
Truss 6	2.07, 0.08	2.14, 0.06	0.092	4.31, 0.16	4.37, 0.18	0.441
Truss 7	3.03, 0.09	3.22, 0.18	0.007 ∗	5.83, 0.27	5.68, 0.31	0.171
Truss 8	2.40, 0.18	2.30, 0.23	0.261	4.26, 0.21	4.27, 0.21	0.931
Truss 9	0.45, 0.03	0.44, 0.03	0.170	2.25, 0.11	2.34, 0.06	0.030 ∗

Fig. 2 Upper: Positions of the nine truss elements used to determine potential differences in morphometry between laboratory-reared lean and siscowet lake trout. Elements were the same as those used by Moore & Bronte (2001) to distinguish wild lean and siscowet lake trout. Measurements were 1, posterior end of maxillary to anterior tip of snout; 2, posterior aspect of neurocranium to origin of pectoral fin; 3, origin of dorsal fin to origin of pelvic fin; 4, origin of pelvic fin to origin of anal fin; 5, origin of anal fin to origin of adipose fin; 6, origin of adipose fin to posterior end of anal fin; 7, origin of adipose fin to anterior attachment of ventral membrane of caudal fin; 8, insertion of anal fin to attachment of dorsal membrane of caudal fin; 9, anterior attachment of ventral membrane of caudal fin to distal margin of caudal peduncle. Lower: Average and standard deviation (SD) for truss measurements (as shown in upper figure) at years 1 and 2 for laboratory-reared lean and siscowet lake trout and P-values calculated by multivariate analysis of variance (MANOVA). Asterisks indicate significant difference between lean and siscowet lake trout at $P \leq 0.05$.

higher gene frequencies in siscowet vs. lean lake trout (Table 2). Similar to the BLASTN-based analysis, there were a few significant comparisons but in the opposite direction from that predicted by gene frequency. Of these, the FK506-binding protein 5 was significant in both analyses (Tables 1 and 2).

Since many of the clusters or assemblies analysed by gene frequency contained a low total number of sequences (Tables 1 and 2), we also used the RPKM method (Mortazavi *et al.* 2008) for determining differences in expression values that takes into consideration relative library size and the overall number of gene reads within contigs (Genomic Workbench, CLCBio). Using this approach, 18 genes with the largest expression value differences across libraries were examined (Table 3). Given the way in which the expression values are calculated, these contigs contained some of the largest number of reads within each library. While these transcripts also exhibited differences in gene frequencies between libraries, the differences were lower than those observed with the two gene frequency analyses discussed above. Of the 18 comparisons showing higher expression values in lean lake trout, 6 were corroborated by qPCR and 17 of the expression differences measured by qPCR were in the correct direction (i.e. lean > siscowet) regardless of the *P*-value (Table 3). Most of the qPCR-corroborated genes also contained the highest frequency differences between gene reads (Table 3). In contrast to the gene frequency analyses, four genes (type-4 ice-structuring pro-

tein LS-12 precursor, acyl-CoA binding protein, similar to neurotoxin/CD59/Ly6, and haptoglobin), that showed higher expression levels in siscowet vs. lean lake trout, were corroborated by qPCR ('similar to neurotoxin/CD59/Ly6' at $P = 0.085$).

Discussion

While past studies have reared lean and siscowet lake trout under hatchery conditions (Eschmeyer & Phillips 1965; Stauffer & Peck 1981), this is the first comprehensive study that quantified a wide range of phenotypic differences between these lake trout forms within the same individuals reared under identical environmental conditions. It is also the first study to use pyrosequencing to quantify differences in transcriptome expression between different ecotypes. The results clearly demonstrate that key phenotypic differences that have been observed between wild lean and siscowet lake trout such as condition factor, morphometry and lipid levels, persist in these two forms when reared in the laboratory under identical environmental conditions. The results strongly suggest that these differences are genetic and not a result of plasticity due to environment.

In differentiating general morphology among lean, siscowet and humper lake trout, Moore & Bronte (2001) found that the 9 (out of 31) truss elements used in the current study were able to discriminate between lean and siscowet lake trout at a level of 80%. Based on
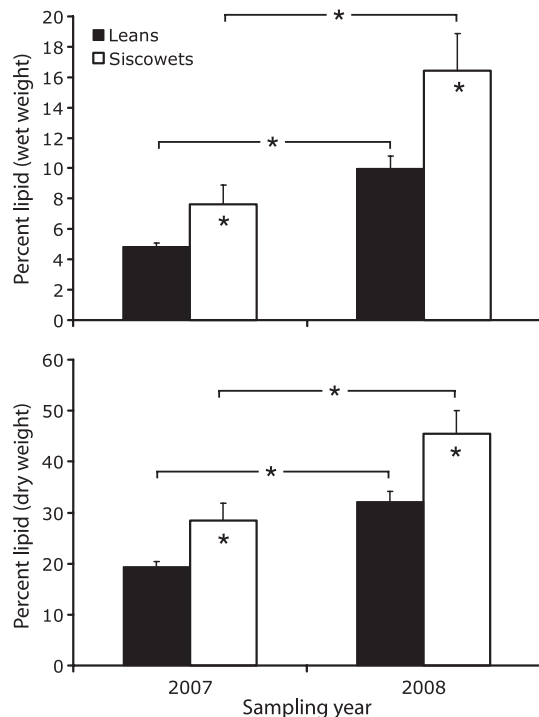
Fig. 3 Mean % lipid + SEM calculated on the basis of wet weight (upper) and dry weight (lower) for $N = 6$ (2007) and $N = 20$ (2008) individuals/lean or siscowet lake trout. Asterisks on bars indicate significant ($P \leq 0.05$) differences between lean and siscowet lake trout per sampling period and on brackets indicate significant ($P \leq 0.05$) differences from year 1 to year 2 within a lake trout form.

these nine elements, we found significant differences between our communally reared siscowet and lean lake trout in three elements at year 1 and six at year 2. Of these, two elements (1 and 3) associated with depth of the body and length of the jaw, were significantly different in both years. Khan & Qadri (1970), applying more classical morphometric techniques, found that siscowet had smaller heads and shorter jaws than lean lake trout. This is consistent with the shorter length of truss element 1, a measurement of the length of the jaw, that we observed in laboratory-reared siscowet in both years. Similarly, truss element 3 is associated with the depth of the body and this was significantly longer in siscowet as compared to lean lake trout in both years. The difference in this truss element between the lake trout forms, correlates well with the higher condition factor of siscowet compared to lean lake trout. In fact, differences in truss element 3 were the most significant ($P = 0.006$ in year 1 and $P = 0.002$ in year 2) of all the truss measurements. Wild siscowet lake trout have longer and thicker caudal peduncles than other lake trout forms (Moore & Bronte 2001), and this correlates with the significant increase we observed in truss element 9 in siscowet lake trout in year 2.

The number of significantly different truss elements was less for our laboratory fish compared to wild lean and siscowet lake trout (Moore & Bronte 2001). Sympatric populations of Arctic char reared under common conditions, maintained significant morphological differences, but the differences were not as great as those seen in wild fish, indicating some phenotypic plasticity (Adams & Huntingford 2004). The lean and siscowet lake trout reared in this study may also be exhibiting some similar plasticity. Another factor may be the size and age of the fish analysed in this study compared to the wild fish analysed by Moore & Bronte (2001). The nine truss elements used in the present study were found to be informative in differentiating wild lean and siscowet lake trout that were 300–900 mm (Moore & Bronte 2001). Fish in the current study averaged 200 and 390 mm in years 1 and 2, respectively. Truss landmarks are more difficult to determine in smaller fish and the size difference between the fish used in each study may have influenced our ability to detect significant differences in trusses on fish assayed in year 1 of the study. In addition, the changes in body morphometry may simply increase with size which is supported by our results showing a greater number of significant truss elements in year 2 than in year 1. Thus, we hypothesize that differences will become greater as the fish become larger and that other truss elements will become significant.

This study found significant differences in weight–length relationships between lean and siscowet lake trout within 1 year of being grown under identical conditions. Laboratory-reared siscowet lake trout were significantly heavier and longer than lean lake trout by 2 years, but the difference in weight was more dramatic. Thus, weight–length growth model parameters were significantly different between laboratory siscowet and lean lake trout ($P < 0.005$; GLM procedure in R version 2.8.1; Fig. 5). For any given length, laboratory-reared siscowet had higher weights than lean lake trout, and this relationship is also observed between wild siscowet and lean lake trout populations in Lake Superior (Fig. 5).
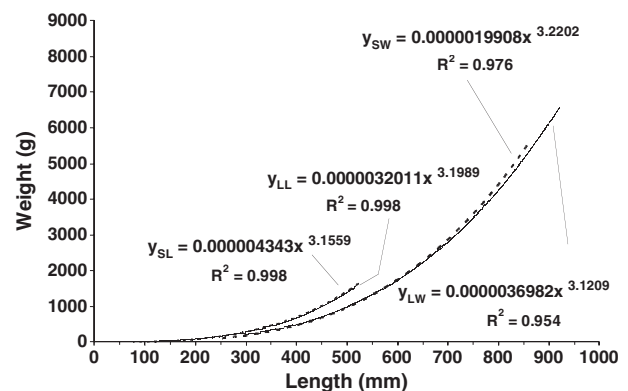
We found significantly higher muscle lipid content in siscowet vs. lean lake trout within 1 year of growth under identical environmental conditions, similar to results reported by Eschmeyer & Phillips (1965) for pond-reared siscowet and lean lake trout. Wild siscowet have a higher lipid level than lean lake trout (Eschmeyer & Phillips 1965; Wang *et al.* 1990) and this difference increases with size (Eschmeyer & Phillips 1965). It is difficult to absolutely compare the lipid levels reported for past studies on wild lake trout to our results since there were differences in the body region sampled and in the lipid analysis itself. However, on a dry weight basis, the laboratory-reared siscowet in our study appeared to have very similar levels in relation to length as reported for
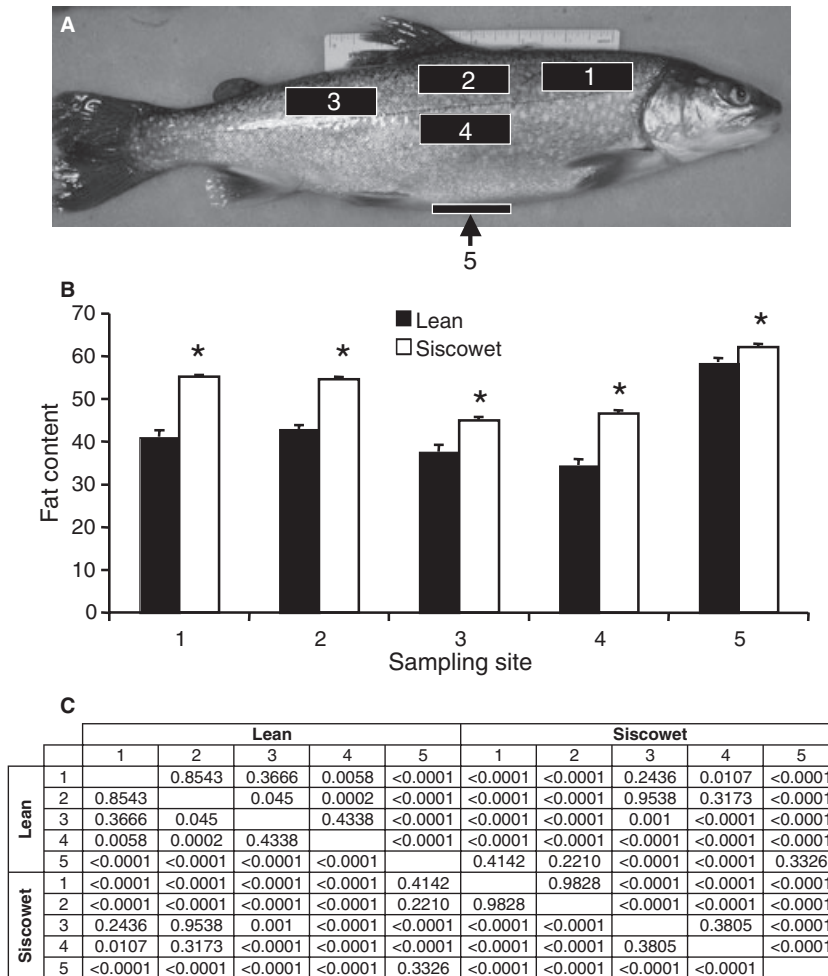
Fig. 4 (A) Regions of the body sampled by the fatmeter. *Note*: the fatmeter was placed on the ventral body wall facing dorsal for region 5. All other regions were on the lateral surface of the right side of the body. (B) Mean fatmeter readings + SEM for each region of the body/lean or siscowet lake trout ($N = 19$). Asterisks indicate significant ($P \leq 0.05$) differences between lean and siscowet lake trout means/sampling region calculated on arcsin transformed data. (C) Matrix of *P*-values for mean comparisons between the fatmeter measurements made on all regions (1–5) of the body per lean and siscowet lake trout. *P*-values ≤0.05 are significantly different.

C

		Lean					Siscowet				
		1	2	3	4	5	1	2	3	4	5
Lean	1		0.8543	0.3666	0.0058	<0.0001	<0.0001	<0.0001	0.2436	0.0107	<0.0001
	2	0.8543		0.045	0.0002	<0.0001	<0.0001	<0.0001	0.9538	0.3173	<0.0001
	3	0.3666	0.045		0.4338	<0.0001	<0.0001	<0.0001	0.001	<0.0001	<0.0001
	4	0.0058	0.0002	0.4338		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	5	<0.0001	<0.0001	<0.0001	<0.0001		0.4142	0.2210	<0.0001	<0.0001	0.3326
Siscowet	1	<0.0001	<0.0001	<0.0001	<0.0001	0.4142		0.9828	<0.0001	<0.0001	<0.0001
	2	<0.0001	<0.0001	<0.0001	<0.0001	0.2210	0.9828		<0.0001	<0.0001	<0.0001
	3	0.2436	0.9538	0.001	<0.0001	<0.0001	<0.0001	<0.0001		0.3805	<0.0001
	4	0.0107	0.3173	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.3805		<0.0001
	5	<0.0001	<0.0001	<0.0001	<0.0001	0.3326	<0.0001	<0.0001	<0.0001	<0.0001	

wild siscowet lake trout (Eschmeyer & Phillips 1965). Our laboratory-reared lean lake trout, however, appear to have higher lipid content than wild lean lake trout at similar sizes. For example, a wild lean lake trout approximately 40 cm long had a lipid level of approximately 20% (Eschmeyer & Phillips 1965), whereas lipid levels in our cultured lean lake trout at the same size were approximately 32% (Fig. 4). Thus, while the difference in lipid levels observed between lean and siscowet lake trout undoubtedly has some genetic basis, other factors such as diet and activity could also influence the lipid levels both between wild lake trout forms and also between the wild and laboratory-reared fish.

Although the techniques used with the fatmeter and the chemical lipid measurements were very different, the general conclusion drawn from the results of these two approaches was the same; siscowet had a higher lipid level in the skeletal muscle than lean lake trout. We did not perform any direct calibration of the fatmeter with actual lipid levels in muscle, so our measures are relative differences between lean and siscowet lake trout. Fatmeter measurements made on the two sam-

$y_{SW} = 0.0000019908x^{3.2202}$
$R^2 = 0.976$

$y_{LL} = 0.0000032011x^{3.1989}$
$R^2 = 0.998$

$y_{SL} = 0.000004343x^{3.1559}$
$R^2 = 0.998$

$y_{LW} = 0.0000036982x^{3.1209}$
$R^2 = 0.954$

Fig. 5 Weight–length relationships for laboratory (subscript L) and wild (subscript W) siscowet (subscript S, dashed lines) and lean (subscript L, solid lines) lake trout from Lake Superior populations. Data for wild fish were collected from lake trout surveys conducted in Michigan waters of Lake Superior near Marquette during 2003–2006. Data presented for laboratory lean and siscowet lake trout are reanalysed from that presented in Fig. 1.

pling positions over the epaxial muscle mass posterior to the head (positions 1 and 2) were higher in both lake trout forms compared to more posterior or ventral positions on the side of the body (Fig. 4). In Pacific salmon, these two regions of the body gave the strongest predictive relationships between gross energy density and fatmeter readings (Crossin & Hinch 2005). In the present study, the highest readings were on the ventral aspect of the body (position 5). However, the body wall is very thin at this point and these readings may be measuring fat that invests the pyloric caeca.

While the adaptive significance of higher fat content in siscowet lake trout is unclear, one hypothesis is that increased lipid is associated with the bioenergetics of vertical migration. Lipid may decrease the costs of maintaining neutral buoyancy at a range of depths (Eshenroder & Burnham-Curtis 1999; Henderson & Anderson 2002) and facilitate vertical and horizontal migration during feeding.

Overall, our study results suggest that there are strong genetic components to the phenotypic differences observed between siscowet and lean lake trout in the wild. Differences in the morphology of various anatomical regions of the rainbow trout body derived from several distinct habitats were also shown to have a strong genetic component in garden variety experiments (Keeley *et al.* 2007). These differences correlated with the habitat (e.g. streams vs. lakes) or feeding (e.g. piscivores vs. nonpiscivores) characteristics of the various rainbow trout ecotypes. In Arctic char, there are many examples of sympatric morphs occurring in postglacial lakes (review Jonsson & Jonsson 2001). Results from various laboratory rearing studies have implicated both genetic and environmental explanations for the differences observed in wild Arctic char populations (Nordeng 1983; Svedang 1990; Hindar & Jonsson 1993; Skulason *et al.* 1996; Klemetsen *et al.* 2002; Adams & Huntingford 2004); however, the degree that genetics or the environment are involved appears to vary among studies. It has been hypothesized that species divergence may initially involve the environmental regulation of alternative discrete phenotypes followed by genetic control (Skulason *et al.* 1999). If so, the degree that a phenotypic trait in Arctic char is explained by genetic or environmental control in rearing experiments, may be related to the level of diversification in the morphs that are being tested (Adams & Huntingford 2004). Regardless, the typical habitats in which sympatric Arctic char morphs are found are deep lakes that have profundal, pelagic and littoral habitats. Most of the lacustrine Arctic char morphs are separated on the basis of habitat and diet; profundal/zoobenthic feeding and limnetic/zooplankton feeding, though other differences may exist between morphs (e.g. spawning time, coloration, body size—Hindar & Jonsson

1982). Differences, particularly in head and mouth structure appear to be related to feeding. For example, pelagic morphs have terminal mouths, long and dense gill rakers and short pectoral fins compared with benthos feeding morphs (Hindar & Jonsson 1982, 1993).

As with Arctic char, deep postglacial lakes also contain lake trout morphs that may be specialized to profundal and pelagic habitats such as seen in Lake Superior with lean and siscowet lake trout. It has been suggested that the longer and thicker caudal peduncle of siscowet lake trout is adaptive to foraging in deeper water where vertical migrations may be extensive and burst swimming necessary for feeding (Moore & Bronte 2001). Similarly, in other large lakes in North America, deepwater and shallow lake trout forms have been reported (Blackie *et al.* 2003; Alfonso 2004; Zimmerman *et al.* 2006, 2007) that are differentiated by body shape (deep vs. elongate), length of the pectoral fins (long vs. short) and buoyancy (high vs. low) (Zimmerman *et al.* 2006). Differences in these characters are also thought to be related to the energetics of swimming and movement in the water column (Zimmerman *et al.* 2006). Besides salmonids, there are examples of fish from other taxa in which sympatric populations within a species demonstrate morphological specializations correlated with feeding in specific habitats. For example, studies have shown that perch in the littoral zone had a significantly deeper body in comparison to pelagic perch and in the laboratory, had higher capture rates in feeding trials in vegetation as opposed to open water (Svanback & Eklov 2003). The opposite was true for pelagic perch that were significantly more streamlined and had higher capture rates in the laboratory in open water trials as compared to trials in vegetation. Thus, differences in morphometry and physiology (i.e. lipid) between sympatric lake trout forms living in divergent habitats are probably related to feeding efficiencies within a specific habitat. As with Arctic char, it is possible that there are other differences between sympatric lake trout forms that are genetically fixed since artificial rearing experiments with nonoverlapping lake trout ecotypes taken from several lakes, have demonstrated a genetic component to egg size and age and size at sexual maturity (McDermid *et al.* 2007).

In the present study we used Roche 454 sequencing technology to examine differences in transcriptome expression between the livers of lean and siscowet lake trout reared under identical environmental conditions. In a study on yellow perch, we used traditional cDNA library construction and capillary sequencing (ABI 3730) to investigate differences in transcript expression in the livers of yellow perch treated with estradiol-17b in the diet (Goetz *et al.* 2009). That study identified 28 oestradiol-regulated (75% difference between libraries) genes. Of 17 of the 28 genes that we examined with qPCR, 14

were confirmed to be regulated by oestrogen. Since very small numbers of sequences were analysed (∼3500/library), we considered that study a successful prelude to the use of pyrosequencing for quantitative transcriptomics through the comparison of gene expression on a much larger scale. Results presented here clearly indicate that sequencing can successfully identify novel genes and delineate differentially expressed transcripts. However, discrepancies between RNA-seq analysis and qPCR are evident. When we started this work, readily available software programs (e.g. CAP3) were incapable of handling the data produced with 454 sequencing on unnormalized cDNA. Thus we initially developed an indirect method to look at gene frequencies by aligning all sequences (from both MID libraries) at the nucleotide level using BLASTN and then grouping sequences based on BLASTN scores. In the study on yellow perch livers, we compared the alignment approach with an analysis of assembled contigs using CAP3. We observed nearly the same transcript frequency results with both bioinformatic approaches (Goetz *et al.* 2009), suggesting that the alignment approach would work with much larger numbers of sequences. However, other algorithms and software packages (e.g. Genomic Workbench, CLCBio) have recently been developed to assemble sequences and examine expression patterns from pyrosequencing. Clearly both bioinformatic approaches were able to delineate transcripts that are constitutively expressed at higher levels in lean vs. siscowet lake trout livers. Further, many of the same genes were identified in both frequency analyses, particularly the genes that could be corroborated by qPCR. However, delineating genes that were expressed at higher levels in siscowet vs. lean lake trout livers was not possible using either of these gene frequency analyses.

The reason for this is unclear but could be a result of several things. First, an inherent problem with library construction was the large amount of cDNA required. We produced 10 μg of cDNA for each library that was constructed and this required SMART™ (Clontech) technology that uses PCR amplification. The amplification may not have been equal across transcripts, resulting in higher levels of certain genes in a given library when the expression of those genes was not really different in the original tissues. The amplification processes used to produce the libraries for GS-FLX sequencing may have further accentuated differences.

When examining qPCR results, we also noticed that expression for some genes was not continuous across individuals, but was either very high or very low, differing by several orders of magnitude. Thus, for these genes, cDNA produced from the pooled livers of one lake trout form could have very large quantities of the transcript if several of these individuals were included in the pool, even though there would not be a significant mean difference in the level of the transcript when assayed by qPCR and averaged over a number of individuals per lake trout form. In addition, when MID libraries are sequenced within a single 454 run, there can be an unequal number of sequences produced between the libraries even if preliminary titering is done before the final sequencing. In the current study, the final ratio of sequences between the two MID libraries was 1:1.75 (siscowet:lean). We corrected gene frequencies by this factor (Tables 1–3), but the unequal number of sequences between the libraries may still have had an impact that is not fully realized.

There could also be problems with the qPCR used to corroborate the RNA-seq results. For example, it has recently been reported that sequence polymorphisms in oysters could lead to large differences in the efficiencies of qPCR reactions between individuals as a result of differences in primer annealing across samples (Taris *et al.* 2008). Given the combined results of the current study it is possible that genetic differences exist between lean and siscowet lake trout that could be present at primer annealing locations. Further, if there are multiple forms of a gene with similar sequences within the primer regions (e.g. superfamily members, duplicate genes), a single qPCR may amplify very similar genes that are not differentially expressed and would, therefore, mask the results of the differentially expressed gene. However, while problems with qPCR might impact the results observed for isolated genes, they are unlikely to be the basis for the inability to corroborate upregulation of genes in siscowet lake trout in general.

Since discrepancies were identified between gene frequency and qPCR in regard to genes expressed at higher levels in the livers of siscowet vs. lean lake trout, we examined the RNA-seq data based on differences in expression values as determined by the RPKM method (Mortazavi *et al.* 2008). This expression analysis accounts for the number of sequences within an assembled contig and for differences in the number of total reads per library. Large expression values, or even large differences in the expression values, do not necessarily translate to large fold differences in gene frequencies between the libraries (see adjusted frequencies, Table 3). However, we hypothesized that given the large number of reads being analysed overall (>250 000), it might be easier to observe a difference in gene expression by chance alone if small numbers of gene reads were involved rather than large numbers of reads. In addition, problems arising from biased cDNA synthesis might impact small numbers of gene reads more so than large ones. Because of the total number of sequence reads involved, the genes from the expression analysis that were analysed were not the same as those

in the gene frequency analyses. Exceptions were the C1q complement protein and the peroxisome proliferator-activated receptor (PPAR) that were present in high copy even in the gene frequency analyses (Table 2). Looking at the contigs containing the greatest differences in expression values, we observed several genes that were expressed at higher levels in lean vs. siscowet lake trout livers that were corroborated by qPCR. Interestingly, most of these genes also contained the highest frequency differences between gene reads (Table 3) even if they were not as great as those in the gene frequency analysis. However, more importantly, we identified four genes that were expressed at higher levels in the livers of siscowet as compared to lean lake trout, and these could be corroborated by qPCR. These were the only transcripts observed to be higher in siscowet lake trout livers throughout all of the analyses.

Clearly, a consideration that must be made in using pyrosequencing for comparative transcriptomic analysis is the way in which samples are prepared for sequencing. Specifically, the use of amplification for cDNA production will probably bias some gene frequencies. Thus, some other method should be used to obtain large amounts of cDNA for 454 sequencing. Alternatively, for nonmodel organisms without characterized genomes, other sequencing technologies (i.e. Illumina Genome Analyzer, ABI SOLiD) could be used for specifically determining gene frequencies against an existing EST backbone, or a backbone developed by 454 sequencing of appropriate samples.

Complementary DNA microarrays have been used to determine the differences in transcripts constitutively expressed in liver and muscle between sympatric whitefish (*Coregonus clupeaformis*) ecotypes (Derome & Bernatchez 2006; Derome *et al.* 2006; St-Cyr *et al.* 2008). Genes that are involved in energy production and muscular activity differed between ecotypes and generally reflected their life histories. For example, dwarf whitefish that occupy the pelagic zone and are active swimmers, had higher expression of genes involved in energy production and muscle contractility compared with normal whitefish that forage on benthic prey and are less active (Derome *et al.* 2006). Lean and siscowet lake trout occupy different habitats; siscowet being deepwater forms while lean lake trout live at shallower depths. As a result of the habitats they are found in, there might be some differences in the overall activity of leans and siscowets as well (e.g. Henderson & Anderson 2002; Hrabik *et al.* 2006) though differential expression of genes related to metabolism were not readily apparent in the current study. Given the higher lipid levels in siscowet vs. lean lake trout, we expected to see differences in genes related to lipid production, metabolism or transport. From the gene frequency and expression analysis we did observe

several lipid-related genes including acyl-CoA desaturase, PPAR gamma (PPARγ), and apolipoprotein B to be expressed at higher levels in lean vs. siscowet lake trout livers. Acyl-CoA desaturase (also called stearoyl-CoA desaturase) is a pivotal enzyme that catalyses the initial oxidation reaction for the desaturation of long-chain saturated fatty acids into monounsaturated fatty acids (Nakamura & Nara 2004). This gene has been studied in several fish species (Hsieh *et al.* 2004) and the activity has been associated with membrane fluidity because of the different melting temperatures of the monounsaturated products formed by the action of the enzyme (Tocher 2003). In mammals, the product of acyl-CoA desaturase, oleic acid, is the major fatty acid of adipose triglycerides (Kokatnur *et al.* 1979) and could, therefore, implicate this enzyme in lipid storage. PPARs are nuclear transcriptional factors that bind fatty acids and eicosanoids (Willson *et al.* 2000). The contig observed in this study aligned (weakly) at the nucleotide level with Atlantic salmon PPARγ. Salmon PPARγ is expressed in the liver and thought to be involved in peroxisomal β-oxidation of liver fatty acids and its presence in the liver was speculated to be the result of lipid deposition (Ruyter *et al.* 1997). Elevated acyl-CoA desaturase and PPARγ in lean lake trout could be related to a preferential storage of lipid in the liver as compared to the muscle in siscowet lake trout. Lipid-related genes that were higher in siscowet vs. lean lake trout livers included acyl-CoA binding protein; a highly conserved protein that binds long-chain acyl-CoA esters and acts as an acyl-CoA transporter (Burton *et al.* 2005; Faergeman *et al.* 2007). The precise role is unknown but in yeast it is involved in fatty acid chain elongation and sphingolipid synthesis (Faergeman *et al.* 2007). Another gene that was higher in siscowet vs. lean lake trout livers was annotated as 'Type-4 ice-structuring protein LS-12 precursor' from Atlantic salmon (accession no. ACI68824). This protein contains an apolipoprotein AII (apoAII) region. Apolipoproteins bind lipids and are fundamental to the packaging of lipids into lipoproteins for transport through the primary circulatory system in animals. In mammals (Schonfeld *et al.* 1978) and fish (Babin & Vernieer 1989), apoAII is the second most abundant protein component of high density lipoproteins. ApoAII is associated with increased levels of plasma fatty acids and triglycerides (as summarized in Castellani *et al.* 2008), and transgenic mice that overexpress apoAII have reduced skeletal muscle fatty acid oxidation and increased triglyceride accumulation (Castellani *et al.* 2001, 2004). Assuming a similar relationship for apoAII in fish, increased expression of this gene may be associated with decreased fatty acid utilization and increased lipid storage in the siscowet muscle.

In addition to genes related to lipids, we observed a significant number of immune-related genes to be

differentially expressed between lean and siscowet lake trout and this was unexpected. In particular, complement component C3, proteasome, FK506 binding protein 5 (immunophilin family member), and several C1q proteins were constitutively expressed at higher levels in lean vs. siscowet lake trout and this was corroborated by qPCR. The gene that was most consistently (across all bioinformatic analyses) and differentially expressed between the two lake trout forms, was the C1q complement protein. In fact, there appeared to be several forms of C1q (C1q-like adipose specific protein, ovary-specific C1q-like factor, complement C1q-like protein 2) that were differentially expressed; some to a greater extent than others based on qPCR (Tables 1–3). C1q is the target recognition protein of the classic complement pathway (Kishore & Reid 2000) and a member of a large family of proteins that contain the C1q domain (Ghai *et al.* 2007) including precerebellin that was also differentially expressed between siscowet and lean lake trout (Table 3). Because of the unique structure of the molecule, C1q can bind to a number of ligands including LPS, porins, phospholipids, IgG, IgM and DNA to initiate a response (Ghai *et al.* 2007; Sjoberg *et al.* 2009). Thus, it plays a pivotal role in complement activation. C1q and other C1q domain containing proteins have been found in other vertebrates including fish (Mei & Gui 2008). Complement component C3 is also a pivotal member of the complement system being at the convergence of all three complement activation pathways: classical, alternative and lectin (Carroll 2004). Interestingly, C1q and C3 have been demonstrated in microarray experiments to be upregulated in the liver during bacterial challenge in channel catfish (*Ictalurus punctatus*) (Peatman *et al.* 2007) and a putative C1q homolog was upregulated in rainbow trout (*Oncorhynchus mykiss*) following stimulation with bacteria (Gerwick *et al.* 2007). Further, these microarray studies also demonstrated the upregulation of several other genes including haptoglobin (catfish and trout), neurotoxin/C59/Ly6-like protein, and catechol-*O*-methyltransferase domain containing 1 (catfish) that we also observed to be transcribed differentially between lean and siscowet lake trout livers (though not always higher in lean lake trout). In fact, the catechol-*O*-methyltransferase domain containing gene has no prior reported relationship with immunity though it was upregulated 14.8-fold in bacterial-challenged catfish (Peatman *et al.* 2007).

Complement factor proteins circulate in the blood and are recognized as key elements in the innate immune response to pathogens. However, they also enhance the adaptive immune response by indirectly activating B and T cells (Carroll 2004). The primary site for the synthesis of complement proteins is the liver, and in mammals this is thought to be primarily constitutive. However, these factors can be further regulated during infections (Carroll 2004) as reported in catfish and trout (Gerwick *et al.* 2007; Peatman *et al.* 2007). If transcript abundance reflects protein synthesis, could it be that lean lake trout have a higher constitutive level of complement than siscowet lake trout? This would suggest that lean lake trout may be more susceptible to pathogen exposure than siscowet, and constitutive elevation in immune factors such as the complement proteins would be adaptive in defending against pathogens. While it has been proposed that siscowet lake trout undergo vertical migrations for feeding (Henderson & Anderson 2002; Hrabik *et al.* 2006), the frequency of this is unknown. If siscowet lake trout are primarily demersal, living at depths greater than 100 m, then the water temperatures that they experience would be approximately 4 °C throughout the year (Sitar *et al.* 2008). These low temperatures may be less conducive to the survival and propagation of pathogens. In contrast, lean lake trout are more pelagic, shallow water forms that experience a much wider variation of temperatures (Mattes 2004), and possibly a greater number of pathogens including those from nearshore sources.

Conclusion

This study has demonstrated a strong genetic component to the phenotypic differentiation observed between wild lean and siscowet lake trout including differences in growth, morphology and lipid levels. The results on truss analysis and lipid levels suggest that environmental effects could also influence some of these differences. While the basis for these differences at the gene level are still unknown, the transcriptomic analysis presented here suggests that there are various physiological processes that could be different between these lake trout forms and that we know very little about the extent of these differences. Lake Superior is the only Great Lake with remnant multiple forms of wild lake trout that are adapted to different habitats. Recovery programs in the lower Great Lakes have focused primarily on lean lake trout from shallow water, but now are broadening their management options to consider re-introductions of deepwater forms (Bronte *et al.* 2008; Markham *et al.* 2008). Thus, understanding the physiological adaptations to living in deepwater habitats by siscowet lake trout would be very informative to Great Lakes lake trout recovery programs.

Acknowledgements

Conflicts of interest

The authors have no conflict of interest to declare and note that the sponsors of the issue had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Adams CE, Huntingford FA (2004) Incipient speciation driven by phenotypic plasticity? Evidence from sympatric populations of Arctic charr *Biological Journal of the Linnean Society*, **81**, 611–618.

Adams CE, Maitland PS (2006) Arctic charr in Britain and Ireland—15 species or one? *Ecology of Freshwater Fish*, **16**, 20–28.

Adams CE, Fraser D, Wilson AJ *et al.* (2006) Patterns of phenotypic and genetic variability show hidden diversity in Scottish Arctic charr. *Ecology of Freshwater Fish*, **16**, 78–86.

Agassiz L (1850) *Lake Superior: Its Physical Character, Vegetation and Animals*, Gould Kendall and Lincoln, Boston, Massachusetts.

Alfonso NR (2004) Evidence for two morphotypes of lake charr, *Salvelinus namaycush*, from Great Bear Lake, Northwest Territories, Canada. *Environmental Biology of Fishes*, **71**, 21–32.

Babin PJ, Vernieer J-M (1989) Plasma lipoproteins in fish. *Journal of Lipid Research*, **30**, 467–489.

Blackie CT, Weese DJ, Noakes DLG (2003) Evidence for resource polymorphism in the lake charr (*Salvelinus namaycush*) population of Great Bear Lake, Northwest Territories, Canada. *Ecoscience*, **10**, 509–514.

Bronte CR, Moore SA (2007) Morphological variation of siscowet lake trout in Lake Superior. *Transactions of the American Fisheries Society*, **136**, 509–517.

Bronte CR, Sitar SP (2008) Harvest and relative abundance of siscowet lake trout in Michigan waters of Lake Superior, 1929–1961. *Transactions of the American Fisheries Society*, **137**, 916–926.

Bronte CR, Ebener MP, Schreiner DR *et al.* (2003) Fish community change in lake superior, 1970–2000. *Canadian Journal of Fisheries and Aquatic Sciences*, **60**, 1552–1574.

Bronte CR, Krueger CC, Holey ME, Toneys ML, Eshenroder RL, Jonas JL (2008) A guide for the rehabilitation of lake trout in Lake Michigan. *Great Lakes Fishery Comm. Misc. Publ.* 2008-01. http://www.glfc.org/pubs/pub.htm #misc.

Brookstein FL, Chernoff B, Elder RL *et al.* (1985) Morphometrics in evolutionary biology. Special Publication 15, Academy of Natural Sciences of Philadelphia. 277.

Brown EH, Eck GW, Foster NR, Horrall RM, Coberly CE (1981) Historical evidence for discrete stocks of lake trout (*Salvelinus namaycush*) in Lake Michigan. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**, 1747–1758.

Burnham-Curtis M (1993) *Intralacustrine speciation of Salvelinus namaycush in Lake Superior: An investigation of genetic and morphological variation and evolution of lake trout in the Laurentian Great Lakes*. PhD Thesis, University of Michigan.

Burnham-Curtis MK, Smith GR (1994) Osteological evidence of genetic-divergence of lake trout (*Salvelinus namaycush*) in Lake-Superior. *Copeia*, **4**, 843–850.

Burton M, Rose TM, Faergeman NJ, Knudsen J (2005) Evolution of the acyl-CoA binding protein (ACBP). *The Biochemical Journal*, **392**, 299–307.

Carlander KD (1950) . *Handbook of Freshwater Fishery Biology*, WM. C Brown Company, Dubuque, Iowa.

Carroll MC (2004) The complement system in regulation of adaptive immunity. *Nature Immunology*, **5**, 981–986.

Castellani LW, Goto AM, Lusis AJ (2001) Studies with apolipoprotein A-II transgenic mice indicate a role for HDLs in adiposity and insulin resistance. *Diabetes*, **50**, 643–651.

Castellani LW, Gargalovic P, Febbraio M *et al.* (2004) Mechanisms mediating insulin resistance in transgenic mice overexpressing mouse apolipoprotein A-II. *Journal of Lipid Research*, **45**, 2377–2387.

Castellani LW, Nguyen CN, Charugundla S *et al.* (2008) Apolipoprotein AII is a regulator of very low density lipoprotein metabolism and insulin resistance. *The Journal of Biological Chemistry*, **283**, 11633–11644.

Chomcynski P (1993) A reagent for the single-step simultaneous isolation of RNA, DNA and protein from cell and tissue samples. *BioTechniques*, **15**, 532–537.

Chomcynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol extraction. *Analytical Biochemistry*, **162**, 156–159.

Conner DJ, Bronte CR, Selgeby JH, Collins HL (1993) Food of salmonine predators in Lake Superior, 1981–1987. *Great Lakes Fishery Commission Technical Report*, **59**, 1–19.

Crossin GT, Hinch SG (2005) A nonlethal, rapid method for assessing the somatic energy content of migrating adult pacific salmon. *Transactions of the American Fisheries Society*, **134**, 184–191.

Derome N, Bernatchez L (2006) The transcriptomics of ecological convergence between 2 limnetic coregonine fishes (Salmonidae). *Molecular Biology and Evolution*, **23**, 2370–2378.

Derome N, Duchesne P, Bernatchez L (2006) Parallelism in gene transcription among sympatric lake whitefish (*Coregonus clupeaformis* Mitchill) ecotypes. *Molecular Ecology*, **15**, 1239–1249.

Eschmeyer PH, Phillips AM (1965) Fat content of the flesh of siscowets and lake trout from Lake Superior. *Transactions of the American Fisheries Society*, **94**, 62–74.

Eshenroder RL (2008) Differentiation of deep-water lake charr *Salvelinus namaycush* in North American lakes. *Environmental Biology of Fishes*, **83**, 77–90.

Eshenroder RL, Burnham-Curtis MK (1999) Species succession and sustainability of the Great Lakes Fish Community. In: *Great Lakes Fisheries Policy and Managemen* (eds Taylor WW, Ferreri CP), Michigan State University Press, East Lansing, Michigan.

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, **8**, 186–194.

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, **8**, 175–185.

Faergeman NJ, Wadum M, Feddersen S *et al.* (2007) Acyl-CoA binding proteins; structural and functional conservation over 2000 MYA. *Molecular and Cellular Biochemistry*, **299**, 55–65.

Gerwick L, Corley-Smith G, Bayne CJ (2007) Gene transcript changes in individual rainbow trout livers following an inflammatory stimulus. *Fish & Shellfish Immunology*, **22**, 157–171.

Ghai R, Waters P, Roumenina LT *et al.* (2007) C1q and its growing family. *Immunobiology*, **212**, 253–266.

Goetz F, MacKenzie S (2008) Functional genomics with microarrays in fish biology and fisheries. *Fish and Fisheries*, **9**, 378–395.

Goetz FW, Rise ML, Rise M *et al.* (2009) Stimulation of growth and changes in the hepatic transcriptome by estradiol-17beta in the yellow perch (*Perca flavescens*). *Physiological Genomics*, **38**, 261–280.

Goodier JL (1981) Native lake trout (*Salvelinus namaycush*) stocks in the Canadian waters of Lake Superior prior to 1988. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**, 1724–1737.

Hansen MJ (1999) Lake trout in the Great Lakes: basinwide collapse and binational restoration. In:*Great Lakes Policy and Management: A Binational Perspectiv* (eds Ferreri CP, Taylor WW). pp. 417–453, Michigan State University Press, East Lansing, Michigan.

Harvey CJ, Schram ST, Kitchell JF (2003) Trophic relationships among lean and siscowet lake trout in Lake Superior. *Transactions of the American Fisheries Society*, **132**, 219–228.

Henderson BA, Anderson DM (2002) Phenotypic differences in buoyancy and energetics of lean and siscowet lake charr in Lake Superior. *Environmental Biology of Fishes*, **64**, 203–209.

Hindar K, Jonsson B (1982) Habitat and food segregation of dwarf and normal arctic charr (*Salvelinus alpinus*) from Vangsvatnet Lake, western Norway. *Canadian Journal of Fisheries and Aquatic Sciences*, **39**, 1030–1045.

Hindar K, Jonsson B (1993) Ecological polymorphism in Arctic charr. *Biological Journal of the Linnean Society*, **48**, 63–74.

Hrabik TR, Jensen OP, Martell SJD, Walters CJ, Kitchell JF (2006) Diel vertical migration in the Lake Superior pelagic community. I. Changes in vertical migration of coregonids in response to varying predation risk. *Canadian Journal of Fisheries and Aquatic Sciences*, **63**, 2286–2295.

Hsieh SL, Chang HT, Wu CH, Kuo CM (2004) Cloning, tissue distribution and hormonal regulation of stearoyl-CoA desaturase in tilapia, *Oreochromis mossambicus*. *Aquaculture*, **230**, 527–546.

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. *Genome Research*, **9**, 868–877.

Jonsson B, Jonsson N (2001) Polymorphism and speciation in Arctic charr. *Journal of Fish Biology*, **58**, 605–638.

Keeley ER, Parkinson EA, Taylor EB (2007) The origins of ecotypic variation of rainbow trout: a test of environmental vs. genetically based differences in morphology. *Journal of Evolutionary Biology*, **20**, 725–736.

Khan NY, Qadri SU (1970) Morphological differences in Lake Superior lake charr. *Journal of the Fisheries Research Board of Canada*, **27**, 161–167.

Kishore U, Reid KB (2000) C1q: structure, function, and receptors. *Immunopharmacology*, **49**, 159–170.

Klemetsen A, Elliott JM, Knudsen R, Sorensen P (2002) Evidence for genetic differences in the offspring of two sympatric morphs of Arctic charr. *Journal of Fish Biology*, **60**, 933–950.

Kokatnur MG, Oalmann MC, Johnson WD, Malcom GT, Strong JP (1979) Fatty acid composition of human adipose tissue from two anatomical sites in a biracial community. *American Journal of Clinical Nutrition*, **32**, 2198–2205.

Krueger CC, Ihssen PE (1995) Review of genetics of lake trout in the great lakes: history, molecular genetics, physiology, strain comparisons, and restoration management. *Journal of Great Lakes Research*, **21**, 348–363.

Lawrie AH, Rahrer JF (1973) Lake Superior: a case history of the lake and its fisheries. In: Great Lakes Fishery Commission Technical Report 19.

Markham JL, Cook A, MacDougall T *et al.* (2008) A strategic plan for the rehabilitation of lake trout in Lake Erie, 2008–2020. In: Great Lakes Fish. Comm. Misc. Publ. 2009-02.

Mattes WP (2004) Temperature and depth profiles of *Salvelinus namaycush* (lake trout) in Lake Superior. In: *Great Lakes Indian Fish and Wildlife Commission Project Report 04-01*, Odanah, Wisconsin.

McDermid JL, Ihssen PE, Sloan WN, Shuter BJ (2007) Genetic and environmental influences on life history traits in lake trout. *Transactions of the American Fisheries Society*, **136**, 1018–1029.

McKinnon JS, Rundle HD (2002) Speciation in nature: the threespine stickleback model systems. *Trends in Ecology & Evolution*, **17**, 480–488.

Mei J, Gui JF (2008) Bioinformatic identification of genes encoding C1q-domain-containing proteins in zebrafish. *Journal of Genetics and Genomics*, **35**, 17–24.

Moore SA, Bronte CR (2001) Delineation of sympatric morphotypes of lake trout in Lake Superior. *Transactions of the American Fisheries Society*, **130**, 1233–1240.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, **5**, 621–628.

Nakamura MT, Nara TY (2004) Structure, function, and dietary regulation of delta6, delta5, and delta9 desaturases. *Annual Review of Nutrition*, **24**, 345–376.

Nordeng H (1983) Solution to the ''Char Problem'' based on the arctic char (*Salvelinus alpinus*) in Norway. *Canadian Journal of Fisheries and Aquatic Sciences*, **40**, 1372–1387.

Page KS, Scribner KT, Burnham-Curtis M (2004) Genetic diversity of wild and hatchery lake trout populations: relevance for management and restoration in the Great Lakes. *Transactions of the American Fisheries Society*, **133**, 674–691.

Peatman E, Baoprasertkul P, Terhune J *et al.* (2007) Expression analysis of the acute phase response in channel catfish (*Ictalurus punctatus*) after infection with a Gram-negative bacterium. *Developmental and Comparative Immunology*, **31**, 1183–1196.

Ray BA, Hrabik TR, Ebener MP *et al.* (2007) Diet and prey selection by Lake Superior lake trout during spring, 1986–2001. *Journal of Great Lakes Research*, **33**, 104–113.

Robinson BW, Wilson DS (1994) Character release and displacement in fishes—a neglected literature. *American Naturalist*, **144**, 596–627.

Ruyter B, Andersen O, Dehli A *et al.* (1997) Peroxisome proliferator activated receptors in Atlantic salmon (*Salmo salar*): effects on PPAR transcription and acyl-CoA oxidase

activity in hepatocytes by peroxisome proliferators and fatty acids. *Biochimica et Biophysica Acta*, **1348**, 331–338.

Schonfeld G, Bailey A, Steelman R (1978) Plasma, apolipoprotein, A-I and A-II levels in hyperlipidemia. *Lipids*, **13**, 951–959.

Sitar SP, Morales HM, Mata MT *et al.* (2008) Survey of siscowet lake trout at their maximum depth in Lake Superior. *Journal of Great Lakes Research*, **34**, 276–286.

Sjoberg AP, Trouw LA, Blom AM (2009) Complement activation and inhibition: a delicate balance. *Trends in Immunology*, **30**, 83–90.

Skulason S, Smith TB (1995) Resource polymorphisms in vertebrates. *Trends in Ecology & Evolution*, **10**, 366–370.

Skulason S, Snorrason SS, Noakes DLG, Ferguson MM (1996) Genetic basis of life history variations among sympatric morphs of Arctic char, *Salvelinus alpinus*. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**, 1807–1813.

Skulason S, Snorrason SS, Jonsson B (1999) Sympatric morphs, populations and speciation in freshwater fish with emphasis on Arctic charr. In:*Evolution of Biological Diversit* (eds Magurran AE, May R). pp. 70–92, Oxford University Press, Oxford.

Stauffer TM, Peck JW (1981) Morphological characteristics of juvenile lake trout of various races. Study Final Report -Michigan Department of Natural Resources Study #408.

St-Cyr J, Derome N, Bernatchez L (2008) The transcriptomics of life-history trade-offs in whitefish species pairs (*Coregonus sp.*). *Molecular Ecology*, **17**, 1850–1870.

Strauss RE, Brookstein FL (1982) The truss: body form reconstruction in morphometrics. *Systematic Zoologist*, **31**, 113–135.

Svanback R, Eklov P (2003) Morphology dependent foraging efficiency in perch: a trade-off for ecological specialization? *Oikos*, **102**, 273–284.

Svedang H (1990) Genetic-basis of life-history variation of dwarf and normal arctic charr, *Salvelinus-alpinus* (L), in Stora-Rosjon, central Sweden. *Journal of Fish Biology*, **36**, 917–932.

Taris N, Lang RP, Camara MD (2008) Sequence polymorphism can produce serious artefacts in real-time PCR assays: hard lessons from Pacific oysters. *BMC Genomics*, **9**, 234.

Taylor EB (1999) Species pairs of north temperate freshwater fishes: evolution, taxonomy, and conservation. *Reviews in Fish Biology and Fisheries*, **9**, 299–324.

Tocher DR (2003) Metabolism and functions of lipids and fatty acids in teleost fish. *Reviews in Fisheries Science*, **11**, 107–184.

Torres TT, Metta M, Ottenwalder B, Schlotterer C (2008) Gene expression profiling by massively parallel sequencing. *Genome Research*, **18**, 172–177.

Wang YJ, Miller LA, Perren M, Addis PB (1990) Omega-3 fatty acids in Lake Superior fish. *Journal of Food Science*, **55**, 71–76.

Willson TM, Brown PJ, Sternbach DD, Henke BR (2000) The PPARs: from orphan receptors to drug discovery. *Journal of Medicinal Chemistry*, **43**, 527–550.

Wilson CC, Mandrak NE (2004) History and evolution of lake trout in shield lakes: past and future challenges. In:*Boreal Shield Watersheds: Lake Trout Ecosystems in a Changing Environmen* (eds Gunn JM, Steedman RJ, Ryder RA), CRC Press, Boca Raton, Florida.

Wilson AJ, Gislason D, Skulason S *et al.* (2004) Population genetic structure of Arctic Charr, *Salvelinus alpinus* from northwest Europe on large and small spatial scales. *Molecular Ecology*, **13**, 1129–1142.

Zhao S, Fernald RD (2005) Comprehensive algorithm for quantitative real-time polymerase chain reaction. *Journal of Computational Biology*, **12**, 1047–1064.

Zimmerman MS, Krueger CC, Eshenroder RL (2006) Phenotypic diversity of lake trout in Great Slave Lake: differences in morphology, buoyancy, and habitat depth. *Transactions of the American Fisheries Society*, **135**, 1056–1067.

Zimmerman MS, Krueger CC, Eshenroder RL (2007) Morphological and ecological differences between shallow- and deep-water lake trout in Lake Mistassini, Quebec. *Journal of Great Lakes Research*, **33**, 156–169.

Supporting Information

Additional supporting information may be found in the online version of this article.

Supplemental Data 1 QPCR primer sequences.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.